



University of Huddersfield Repository

Moore, David J.

The development of a design tool for 5-speaker surround sound decoders

Original Citation

Moore, David J. (2009) The development of a design tool for 5-speaker surround sound decoders. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/9050/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

THE DEVELOPMENT OF A DESIGN TOOL FOR 5-SPEAKER SURROUND SOUND DECODERS

by

John David Moore

A thesis submitted to the University of Huddersfield
in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

University of Huddersfield
Queensgate, Huddersfield, UK

(July, 2009)

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of Copyright works, for example graphs and tables (“Reproductions”), which may described by in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Abstract

This thesis presents the development of a software-based decoder design tool (DDT) for producing Ambisonic decoders optimised for playback over 5-speaker layouts. The research specifically focuses on developing decoders for irregular layouts with loudspeakers at a constant radial distance from the central listening position. It was motivated by the desire to provide better surround sound over the standard ITU 5-speaker layout for listeners in the sweet spot and off-centre positions. A wide-ranging literature review is presented revealing the need for such work.

The DDT employs the Tabu Search algorithm to seek improved decoder parameters according to a multi-objective fitness function. The fitness function encapsulates criteria from psychoacoustic models as a set of objectives. In order to ensure the objectives were treated equally a method known as ‘range-removal’ was used for the first time in Ambisonic decoder design. A companion technique termed ‘importance’ allows the systematic prioritisation of range-removed objectives giving a designer control over desired decoder criteria.

Additional elements exist in the DDT that can be turned on or off in different combinations. They include: a novel component for producing decoders with even performance by angle, a novel component for producing performance that correlates with the pattern of human spatial resolution estimated in previous Minimum Audible Angle experiments, and the ability to produce frequency dependent or independent decoders of different orders. Moreover, the user of the DDT can optimise performance for a single listener or multiple distributed listeners. To make the DDT as interactive as possible searches can optionally run on a High Performance Computer.

This thesis also details the extensive testing of Ambisonic decoders for the ITU layout. Decoders have been assessed subjectively in listening tests and objectively using binaural measurements which has verified the methods developed in this research and the DDT’s concept. Furthermore, decoders derived by the DDT have been compared to existing decoders and the results show they give equal or better performance.

The development of a fully-functioning DDT which incorporates techniques for range-removal, importance, even performance by angle, minimum audible angle, off-centre listeners and their use in any combination represent the key outcomes of this work.

To my wife, Sophie.

Acknowledgements

I would first like to thank my supervisor, Dr Jonathan Wakefield, for his guidance and invaluable input to this research work. He is an excellent teacher with a talent for explaining difficult concepts as well as having exceptional proof reading skills.

Thank you to all others who have contributed to the project. In particular, senior lecturers Dr Bruno Fazenda and Braham Hughes, and my fellow research students Matthew Wankling and Julian Romero-Perez. I would also like to thank senior technician Ben Evans, placement students Nicolas Wilhelm and Matthieu Forest from L'Institute Universitaire de Technologie de Cachan as well as Mark Bokowiec from the School of Music, Humanities and Media for help during the experiments.

I am very grateful to Norman Barrett and his son John for sparking my interest off in music technology and providing me with the necessary skills to build upon at University.

Of course, this work would not have been possible without the love and support of my family. This includes the care of my sister, Emily, and my parents-in-law during the write up stage. I especially thank my wonderful wife, Sophie, for her love and patience while completing this work... I very much appreciate it. Finally, my parents... I am endlessly grateful to my parents, for everything they have done for me.

Table of Contents

Abstract	3
Acknowledgements	5
Table of Contents	6
List of Figures	12
List of Tables	17
Chapter 1	19
1.1 Motivation	19
1.2 Objectives of this work	20
1.3 Overview and structure of the thesis	21
Chapter 2	22
2.1 Introduction	22
2.2 Auditory localisation	22
2.2.1 Early research	23
2.2.2 Interaural Level Difference	23
2.2.3 Interaural Time Difference and the Interaural Phase Difference	24
2.2.4 Head Related Transfer Functions	26
2.2.5 Head movements	30
2.2.6 Localisation accuracy and spatial resolution	30
2.2.7 The precedence effect	31
2.3 Surround sound	33
2.3.1 A historical perspective	33
2.3.2 Industry standard surround sound loudspeaker configurations	35
2.3.3 Typical surround sound loudspeaker configurations	36
2.3.4 General review of surround sound reproduction techniques	38
2.3.4.1 Positioning of sound sources using inter-channel differences	38
2.3.4.2 Reproduction that takes into account the listener's congenital features	41
2.3.4.3 Wavefield reconstruction methods	42
2.3.5 Subjective comparisons between surround sound reproduction techniques	44
2.3.6 Objective measures for evaluating surround sound systems	46
2.3.6.1 Models of auditory localisation	46

2.3.6.2 Soundfield reconstruction analysis	51
2.4 Optimisation using computer search algorithms	52
2.4.1 Search algorithms used for optimisation	56
2.4.1.1 Exhaustive search	56
2.4.1.2 Heuristic searches	56
2.5 High performance computing	59
2.6 Summary	60
Chapter 3	63
3.1 Introduction	63
3.2 Velocity and energy localisation vectors	63
3.3 Ambisonic theory	66
3.3.1 Encoding	66
3.3.2 Decoding	70
3.3.2.1 Decoders for regular loudspeaker arrays	71
3.3.2.2 Decoders for irregular loudspeaker arrays	74
3.3.2.3 Additional decoding considerations	76
3.4 Tabu search	79
3.5 Summary	82
Chapter 4	83
4.1 Introduction	83
4.2 Improved multi-objective fitness function	84
4.2.1 Volume objectives	84
4.2.2 Vector angle objectives	85
4.2.3 Angle match objective	86
4.2.4 Vector magnitude objectives	87
4.2.5 Implementation details	87
4.2.6 Evaluating frequency dependent decoders	88
4.2.7 Summary	88
4.3 Range-Removal and Importance	89
4.3.1 Objective dominance	89
4.3.2 Range-removal	91
4.3.3 Importance	92

4.3.4 Implementation details	92
4.3.5 Summary	93
4.4 Optimisation of higher order decoders.....	93
4.5 Even localisation performance optimisation.....	94
4.5.1 An analysis of a typical first order Ambisonic decoder for the ITU layout.....	95
4.5.2 Even performance design criteria.....	97
4.5.3 Summary	98
4.6 Exploiting human spatial resolution	98
4.6.1 Auditory localisation resolution.....	98
4.6.2 MAA optimisation criteria	100
4.6.3 Summary	101
4.7 Optimisation of decoders for off-centre listeners.....	101
4.7.1 Background	102
4.7.2 Off-centre evaluation criteria	104
4.7.3 Implementation details	106
4.7.4 Summary	107
4.8 Search Acceleration using High Performance Computing Hardware.....	107
4.8.1 Implementation details	108
4.8.2 Summary	110
4.9 Decoder design tool user interface.....	110
4.9.1 Main user interface.....	111
4.9.2 Performance panel.....	111
4.9.3 Options panel	115
4.10 Code testing	116
4.11 Summary	117
Chapter 5.....	118
5.1 Introduction.....	118
5.2 Design tool settings.....	118
5.3 Testing range-removal and importance.....	119
5.4 Evaluation of the improved multi-objective fitness function.....	123
5.5 The generation of higher order decoders	125
5.6 Evaluation of the even performance optimisation component.....	127

5.7 Evaluation of the minimum audible angle optimisation component.....	132
5.8 Evaluation of the off-centre optimisation components.....	133
5.9 Search algorithm acceleration using High Performance Computing hardware.....	141
5.10 Summary	144
Chapter 6	146
6.1 Introduction	146
6.2 Experimental setup	147
6.3 Test subjects	149
6.4 Test 1 - Real sound source localisation	150
6.4.1 Test procedure	150
6.4.2 Results	151
6.4.3 Summary	160
6.5 Test 2 - Decoded sound source localisation from the central listening position	161
6.5.1 Decoders under assessment	161
6.5.2 Test procedure	162
6.5.3 Results	162
6.5.3.1 Front-back reversals	162
6.5.3.2 Preliminary analysis	163
6.5.3.3 Low frequency noise	166
6.5.3.4 Mid/high frequency noise.....	173
6.5.3.5 Male speech.....	178
6.5.3.6 Overall decoder performance	183
6.5.4 Discussion	184
6.6 Test 3 - Decoded sound source localisation from off-centre listening positions.....	187
6.6.1 Decoders under assessment	187
6.6.2 Test procedure	187
6.6.3 Results	189
6.6.3.1 Preliminary analysis	189
6.6.3.2 Low frequency noise	190
6.6.3.3 Mid/high frequency noise.....	196
6.6.3.4 Male speech.....	201
6.6.3.5 Overall decoder performance	206

6.6.4 Discussion	208
6.7 Summary	209
Chapter 7	211
7.1 Introduction.....	211
7.2 Experimental setup.....	211
7.3 Data processing.....	212
7.4 Estimation of the auditory cues using an auditory model	213
7.4.1 Real source results	216
7.5 Test 1 - Central listening position measurements	218
7.5.1 Results.....	218
7.5.1.1 ITD	218
7.5.1.2 ILD	221
7.5.2 Discussion	223
7.6 Test 2 - Off-centre listening position measurements	224
7.6.1 Results.....	224
7.6.1.1 ITD	224
7.6.1.2 ILD	231
7.6.2 Overall decoder performance.....	237
7.6.3 Discussion	240
7.7 Summary	240
Chapter 8.....	241
8.1 Introduction.....	241
8.2 Further optimisation of the Craven decoder	241
8.3 Further optimisation of the Poletti decoder.....	245
8.4 Summary	247
Chapter 9	248
9.1 Introduction.....	248
9.2 Summary of the main contributions of this thesis.....	248
9.3 Conclusions.....	253
Chapter 10.....	254
Appendix A.....	257
Appendix B	258

Appendix C.....	261
Bibliography.....	272

Word Count: 49,049

List of Figures

Figure 2.1: Interaural Level Difference	24
Figure 2.2: Interaural Time Difference	25
Figure 2.3: Interaural Phase Difference.	26
Figure 2.4: The cone of confusion	27
Figure 2.5: Left and right ear HRTFs for 3 subjects at 3 different angles (0°, 45° and 85°).	29
Figure 2.6: Summing localisation, localisation fusion and the echo threshold.....	32
Figure 2.7: The potential problem of the precedence effect in a 2-channel listening situation	33
Figure 2.8: The standard ITU 5.1 loudspeaker arrangement.	35
Figure 2.9: First example of a typical 5.1 setup in a domestic environment	37
Figure 2.10: Second example of a typical 5.1 setup in a domestic environment	37
Figure 2.11: Inter-channel amplitude differences result in ear phase differences.	39
Figure 2.12: A 2D test function known as the Michalewicz function.....	55
Figure 3.1: The Soundfield microphone	67
Figure 3.2: The angular response of the W, X and Y B-format components.....	68
Figure 3.3: First order to forth order encoding functions.....	70
Figure 3.4: A range of first order virtual microphone directivities.....	72
Figure 3.5: Three different types of Ambisonic decoding for a hexagonal loudspeaker array.....	73
Figure 3.6: The performance of a ‘cardioid’ decoder for the ITU 5-speaker array..	74
Figure 3.7: Schematic diagram of a first order dual band decoder for irregular arrays	77
Figure 3.8: The Tabu search algorithm.....	80
Figure 4.1: Software-based decoder design tool.....	83
Figure 4.2: Vector angle match problem	86
Figure 4.3: The performance of a typical first order frequency independent 5-speaker decoder. .	95
Figure 4.4: The distance/angle of each speaker changes according to the listening position.	104
Figure 4.5: Clearspeed x620 board	108
Figure 4.6: An example C ⁿ program for the ClearSpeed HPC hardware.....	109
Figure 4.7: Decoder design tool structure	111
Figure 4.8: Main interface of the decoder design tool	113
Figure 4.9: Performance panel of the design tool	114
Figure 4.10: Search options panel of the design tool	115

Figure 5.1: Performance plot of a first order decoder derived without range-removal.	120
Figure 5.2: Performance plot of a first order decoder derived with range-removal.	121
Figure 5.3: Performance plot of a first order decoder derived with a greater importance given to the mid/high frequency angle objective.	122
Figure 5.4: Virtual microphone response of typical decoders from first to fourth order	126
Figure 5.5: A good fourth order decoder derived using the design tool.....	127
Figure 5.6: Total error by angle for the even error optimised decoders and a typical decoder ...	129
Figure 5.7: Individual objective error by angle for all three even error decoders	129
Figure 5.8: The decoder design tool performance plot for the best even error decoder.....	131
Figure 5.9: The decoder design tool performance plot for the MAA optimised decoder.....	132
Figure 5.10: The off-centre positions that were evaluated in the fitness function.	134
Figure 5.11: Local velocity vectors at each position evaluated in the fitness function.	136
Figure 5.12: Local velocity vectors at each position evaluated in the fitness function	137
Figure 5.13: Local energy vectors at each position evaluated in the fitness function	138
Figure 5.14: Local energy vectors at each position evaluated in the fitness function.	139
Figure 5.15: Mean magnitude/angle error for the velocity and energy vectors at each position.	140
Figure 5.16: Mean total fitness and 95% confidence intervals.....	143
Figure 6.1: RT60 of the music studio used for the listening tests	147
Figure 6.2: Geometry of the loudspeaker array in the listening tests	148
Figure 6.3: The amplitude envelope of the low and mid/high frequency noise stimuli	149
Figure 6.4: User interface of the real source listening test software	150
Figure 6.5: Subject response versus the actual source angle for low frequency noise	151
Figure 6.6: Subject response versus the actual source angle for mid/high frequency noise.....	152
Figure 6.7: Subject response versus the actual source angle for male speech.....	152
Figure 6.8: Multiple comparison test between the 14 subjects for the low frequency noise source. This is for the original listening test data with no front-back reversal correction applied.	156
Figure 6.9: Multiple comparison test between the 14 subjects for the male speech source. This is for the test data with the reversal correction applied.	157
Figure 6.10: Mean localisation error by angle across all listening test subjects for the low frequency noise source with 95% confidence intervals.....	158
Figure 6.11: Mean localisation error by angle across all listening test subjects for the mid-high frequency noise source with 95% confidence intervals.....	159

Figure 6.12: Mean localisation error by angle across all listening test subjects for the male speech source with 95% confidence intervals.	159
Figure 6.13: Percentage of front-back reversals for each decoder and for each sound source. ...	163
Figure 6.14: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the low frequency noise test. This is the original data without the front-back reversal correction applied.....	167
Figure 6.15: Overall mean error with 95% confidence intervals for each decoder from the low frequency noise test (original data).....	169
Figure 6.16: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the low frequency noise test. The data presented in this figure includes the front-back reversal correction.....	171
Figure 6.17: Overall mean error with 95% confidence intervals for each decoder from the low frequency noise test (with the front-back reversal correction).....	172
Figure 6.18: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the mid/high frequency noise test. This is the original data without the front-back reversal correction.....	174
Figure 6.19: Overall mean error with 95% confidence intervals for each decoder from the mid/high frequency noise test (original data).....	175
Figure 6.20: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the mid/high frequency noise test. The data presented in this figure has the front-back reversal correction applied.....	176
Figure 6.21: Overall mean error with 95% confidence intervals for each decoder from the mid/high frequency noise test (with the front-back reversal correction).....	177
Figure 6.22: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the male speech test. This is the original data without the front-back reversal correction.	179
Figure 6.23: Overall mean error with 95% confidence intervals for each decoder from the male speech test (original data).	180
Figure 6.24: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the male speech test. The data presented in this figure has the front-back reversal correction applied.....	182

Figure 6.25: Overall mean error with 95% confidence intervals for each decoder from the male speech test (with the front-back reversal correction).....	183
Figure 6.26: Overall mean localisation error with 95% confidence intervals for each decoder taking into account the three sound source tests (original and reversal corrected).	184
Figure 6.27: Listening positions (P1 to P9) which were evaluated in the off-centre test.....	188
Figure 6.28: Mean response angle versus the actual source angle for each decoder in the low frequency noise test (listening positions 4 and 8).....	194
Figure 6.29: Overall mean errors with 95% confidence intervals for each decoder at each listening position for the low frequency noise source.	195
Figure 6.30: Mean response angle versus the actual source angle for each decoder in the mid/high frequency noise test (listening positions 2 and 6).....	199
Figure 6.31: Overall mean errors with 95% confidence intervals for each decoder at each listening position for the mid/high frequency noise source.	200
Figure 6.32: Mean response angle versus the actual source angle for each decoder in the male speech test (listening positions 2 and 7). Loudspeakers are shown as red squares. Dashed line is the ideal response. Note the loudspeaker bias effect in both plots.....	204
Figure 6.33 : Overall mean errors with 95% confidence intervals for each decoder at each listening position for the male speech source.....	205
Figure 6.34: Overall mean localisation error at each position (with 95% confidence intervals) taking into account the data from all the sound source tests.	207
Figure 6.35: Overall mean error and 95% confidence intervals for each decoder taking into account all positions for each sound source test.	208
Figure 7.1: HRIR for a real source at 0 degrees	212
Figure 7.2: Truncated HRIR for a source at 0 degrees.....	213
Figure 7.3: Structure of the auditory model used for estimating the interaural cues.....	214
Figure 7.4: ITD of the Neumann dummy head calculated using the developed auditory model.	217
Figure 7.5: ILD of the Neumann dummy head calculated using the developed auditory model.	217
Figure 7.6: ITD for each decoder at the centre listening point (blue solid line). The ITD for a real source is included for reference (red dashed line). The error is shown as a black dash-dot line.	219
Figure 7.7: Mean subject response in the listening test for the low frequency noise source. The red dashed line indicates the ideal response.	221
Figure 7.8: ILD for each decoder measured from the centre listening point.....	222

Figure 7.9: ITD by angle for off-centre tested decoders at the centre position.....	227
Figure 7.10: ITD by angle for off-centre tested decoders at the front position.....	227
Figure 7.11: ITD by angle for off-centre tested decoders at the front-left position.....	228
Figure 7.12: ITD by angle for off-centre tested decoders at the left position	228
Figure 7.13: ITD by angle for off-centre tested decoders at the back-left position	229
Figure 7.14: ITD by angle for off-centre tested decoders at the back position.....	229
Figure 7.15: ITD by angle for off-centre tested decoders at the back-right position.....	230
Figure 7.16: ITD by angle for off-centre tested decoders at the right position.....	230
Figure 7.17: ITD by angle for off-centre tested decoders at the front-right position.....	231
Figure 7.18: ILD by angle for off-centre tested decoders at the centre position.....	233
Figure 7.19: ILD by angle for off-centre tested decoders at the front position.....	233
Figure 7.20: ILD by angle for off-centre tested decoders at the front-left position.....	234
Figure 7.21: ILD by angle for off-centre tested decoders at the left position	234
Figure 7.22: ILD by angle for off-centre tested decoders at the back-left position	235
Figure 7.23: ILD by angle for off-centre tested decoders at the back position.....	235
Figure 7.24: ILD by angle for off-centre tested decoders at the back-right position.....	236
Figure 7.25: ILD by angle for off-centre tested decoders at the right position.....	236
Figure 7.26: ILD by angle for off-centre tested decoders at the front-right position.....	237
Figure 8.1: Performance plot of the best solution produced by the design tool when using Craven's decoder as a starting point.....	243
Figure 8.2: Performance plot of the decoder derived by Craven.	243
Figure 8.3: Zoomed in view of the centre and left front virtual microphones for the new decoder derived from Craven's solution and the original Craven solution.	244
Figure 8.4: Mean velocity vector and energy vector errors for the new decoder derived from Poletti's decoder and the original Poletti's decoder at each position evaluated.....	246

List of Tables

Table 4.1: Core multi-objective fitness function algorithm described using pseudo code	88
Table 4.2: Approximate ranges of the improved fitness function objectives. The objectives with the largest ranges (highlighted) are likely to dominate the search.	90
Table 4.3: Improved fitness function algorithm with range-removal and importance.....	93
Table 4.4: The number of decoder coefficients required for decoding over left/right symmetrical 5-speaker layouts (frequency dependent and independent)	94
Table 4.5: Standard deviation of the fitness function objectives values	96
Table 4.6: Stimuli, duration and the number of subjects from a number of MAA experiments ...	99
Table 4.7: Estimated MAA values from the aforementioned experiments (see previous table)..	100
Table 4.8: MAA objective weightings.....	101
Table 4.9: Off-centre fitness function algorithm	106
Table 5.1: Design tool search settings used when deriving the decoders	118
Table 5.2: Fitness function objective values of the best solutions encountered during design tool applications without the range-removal component and with the range-removal component. ...	120
Table 5.3: Fitness function objective values of the best solution produced by the design tool when giving higher importance to the mid/high frequency angle objective.....	121
Table 5.4: Objective values for the best solutions produced when testing the impact of the new objectives added to the fitness function.	124
Table 5.5: Best solutions produced for each decoder order in an equal importance application.	125
Table 5.6: Standard deviation of objective error for all three decoders.....	130
Table 5.7: Even error decoder objective importance weightings.....	131
Table 5.8: Objective importance weightings used when deriving the off-centre decoder	141
Table 5.9: Comparison of the different search versions.	142
Table 6.1: Total number of front-back localisation confusions by source angle for each test.....	153
Table 6.2: Information about the decoders used in the central listening point test.....	161
Table 6.3: Standard deviation of the mean localisation error by angle and the mean difference from the equivalent real source data for the low frequency noise test (original data).	169
Table 6.4: Standard deviation of the mean localisation error by angle and the mean difference from the equivalent real source data for the low frequency noise test (with reversal correction)	172

Table 6.5: Standard deviation of the mean localisation error by angle and the mean difference from the equivalent real source data for the mid/high frequency noise test (original data).....	175
Table 6.6: Standard deviation of the mean localisation error by angle and the difference from the equivalent real source data for the low frequency noise test (with the reversal correction).	177
Table 6.7: Standard deviation of the mean error by angle and the mean difference from the equivalent real source data for the male speech test (original data).....	180
Table 6.8: Standard deviation of the mean error by angle and the mean difference from the equivalent real source data for the male speech test (with the reversal correction).....	183
Table 6.9: Information about the decoder used in the off-centre listening test.....	187
Table 6.10: Mean error by angle for the low frequency noise source - positions 1 to 5	192
Table 6.11: Mean error by angle for the low frequency noise source - position 6 to 9	193
Table 6.12: Mean error by angle for the mid/high frequency noise source - positions 1 to 5	197
Table 6.13: Mean error by angle for the mid/high frequency noise source - positions 6 to 9	198
Table 6.14: Mean error by angle for the male speech source – positions 1 to 5	202
Table 6.15: Mean error by angle for the male speech source - positions 6 to 9.....	203
Table 7.1: Mean unsigned ITD error by angle (ms) for all decoders and all measurement positions. All decoders exhibit a higher error at the rear of the system.	226
Table 7.2: Mean unsigned ILD error by angle (dB) for all decoders and positions.....	232
Table 7.3: Mean ITD error for each decoder at each listening position. The mean and standard deviation are included in the right hand columns	239
Table 7.4: Mean ILD error for each decoder at each listening position. The mean and standard deviation are included in the right hand column.	239
Table 8.1: Importance weights used when deriving a solution from Craven’s decoder	242
Table 8.2: Fitness function objective values for the Craven decoder and the decoder derived using Craven’s decoder coefficients as a starting point.....	242
Table 8.3: Importance weights used when deriving a solution from Poletti’s decoder	245
Table 8.4: Fitness function objective values for the decoder derived from Poletti and the original Poletti decoder.	246

Chapter 1

Introduction

We are surrounded by sound. Sonic events continuously occur around us and our highly refined, evolutionarily developed hearing system decodes it to provide us with a large amount of information: the ability to perceive the direction, distance and size of the source of these events and also about the space they take place in.

Since Thomas Edison invented the mechanical phonograph cylinder in 1877, sound reproduction technology has developed substantially from simple monophonic playback, to elaborate multichannel surround sound systems capable of physically reconstructing a soundfield. Today, systems for the reproduction of surround sound are widely commercially available and many people have them in their homes. The most common systems in use are 5-speaker and 7-speaker systems. It is estimated that over 75 million people own one of these systems (Home Audio Division 2008). This figure will increase further still if considering in-car audio systems, cinemas and studios. Current usage ranges from the playback of music and sound for movies, to the enhancement of sound in computer games.

Now that we have the means to reproduce surround sound so readily available, the ongoing quest in the audio engineering industry is to enhance it further by providing our hearing system with more information. Harnessing the power of recent improvements in digital technology and computer processing power offers new opportunities to achieve this.

1.1 Motivation

The basic aim of any surround sound reproduction system is to generate the illusion of acoustic reality. The ideal scenario would be to create an acoustic world that is indistinguishable from what we normally hear around us. However, this illusion can only accurately be achieved if a number of conditions are met. These include: accurate sound source localisation, creating a

realistic impression of sound source size and form, listener envelopment (the feeling of being surrounded by sound), accurate perception of sound source movement, and accurate perception of sound source distance. Existing systems can meet some of these conditions to a greater or lesser extent at the optimal listening area known as the “sweet spot”. However, for many systems this area is often only large enough to accommodate a very small proportion of potential listeners. Consequently, listeners positioned outside the sweet spot will receive a degraded surround sound experience.

What are needed are algorithms for enhancing the listening experience. These algorithms should not only focus on improving surround sound at the sweet spot, but also aim to improve surround sound at other positions to enable more listeners to experience simultaneously high-quality surround sound playback. Successfully achieving this goal has great potential for increasing the commercial value of any surround sound system.

1.2 Objectives of this work

The primary objective of this research is to develop algorithms that provide improved playback for surround sound over existing commercial surround sound systems. It will mainly consider 5-speaker systems with a constant radial distance from the centre such as ITU 5.1 (ITU 1994) due to their widespread use both commercially and domestically. The algorithms developed in this work will aim to produce surround sound decoders with improved playback at the sweet spot as well as other positions in the listening area through the manipulation of psychological perception.

The other major objective of this research is to create a software-based decoder design tool that encapsulates all of the algorithms developed as part of this work. The tool should be easy to operate and allow users to tailor the performance of surround sound decoders for both standard and non-standard 5-speaker layouts.

The focus of the work presented in this thesis will be on improving sound source localisation as this is a fundamental feature of human hearing. Other perceptual elements which are important

for realistic surround sound reproduction rely to a large extent on the ability of the listener to first localise the general direction of the sound source (e.g. sound source distance perception and sound source movement).

1.3 Overview and structure of the thesis

There follows, in **Chapter Two**, a general review of the literature. This will examine in detail several topics relevant to this research; auditory localisation, surround sound, optimisation using computers and high performance computing. **Chapter Three** will provide the reader with background theory on the methods adopted in this research. **Chapter Four** describes in detail the developed decoder design tool and each of its individual components. **Chapter Five** examines the theoretical localisation performance of several decoders produced by the design tool. These decoders were further evaluated in two practical experiments that are described in **Chapter six** and **Chapter Seven** respectively. **Chapter eight** will describe the optimisation of some existing surround sound decoders. Finally, a summary of the main contributions with conclusions will be given in **Chapter nine**, and future work suggestions will be given in **Chapter Ten**.

Chapter 2

General review of relevant literature

2.1 Introduction

This chapter will discuss and appraise research within the following relevant subject areas:

- i. Human auditory localisation
- ii. Surround sound
- iii. Optimisation using search algorithms
- iv. High-performance computing

Further appraisal of literature will supplement later chapters which focus on the more detailed problems of this research.

2.2 Auditory localisation

Knowledge of the human auditory system is of paramount importance to audio engineers. For example, engineers have made extensive use of psychoacoustic knowledge in the development of audio coding and compression algorithms (Madisetti & Williams 1997; Painter & Spanias 2000; Ville Pulkki 2007). Likewise, engineers use psychoacoustic knowledge in the configuration of sound reinforcement systems (Mapp 2007). When developing spatial sound reproduction systems our perceptual ability must be considered. How do humans perceive sound in space? What cues does the human auditory system require for accurate sound placement? These are the questions which must first be addressed in order to develop spatial sound reproduction systems. The goal of high quality spatial sound reproduction can only be realised if all the auditory cues are correctly reproduced or simulated for the listener(s).

There follows an outline of important early research in psychoacoustics. Subsequently, a detailed description of each auditory cue will be given as well as background on other psychoacoustic attributes such as human spatial resolution and the role of head movement in localising sound.

2.2.1 Early research

At the start of the 20th century Lord Rayleigh – one of the most important scientists of the modern age - investigated the mechanisms we use for perceiving a sound's direction (Rayleigh 1907). His research looked at how sound level differences between sound waves arriving at the ears assist in sound source localisation. During the investigation Lord Rayleigh found that level differences for low frequency sound would not be great enough to be useful for localisation. This was because when a sound's wavelength is larger than the diameter of the head (below about 1400Hz) the sound waves diffract around the head and the head no longer attenuates sound travelling to the furthest ear. Following this observation Lord Rayleigh looked for alternative reasons why subjects in his experiments could localise lateral low frequency sounds with relative ease. He later established that localisation in the low frequency range is based on time differences incurred because of the time taken for the sound to travel the distance separating the ears. From this work Lord Rayleigh proposed one of the first models for explaining sound source localisation using interaural time differences (ITD) and interaural level differences (ILD), which has been labeled the “Duplex Theory”.

Since Lord Rayleigh's work, sound localisation has been studied extensively and it is now firmly established that the major cues for sound localisation are the interaural level difference (ILD), the interaural time difference (ITD), the interaural phase difference (IPD), and the monaural and binaural spectral cues caused by sound interaction with the external ear and the upper body (Blauert 2001; Brian Moore 2003). The following sections will look at each of these in turn.

2.2.2 Interaural Level Difference

For frequencies above about 1.5 kHz level differences between sound arriving at the ears are used to locate a sound source. Above this frequency a sound's wavelength is shorter than the diameter of the listener's head causing sound waves to be attenuated on their route to the furthest ear (see figure 2-1).

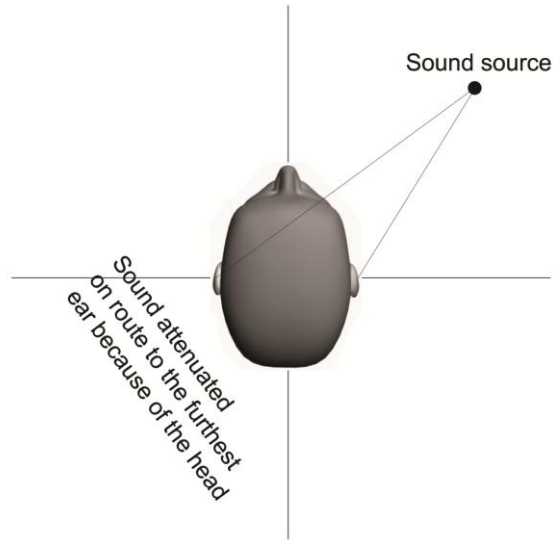


Figure 2-1: Interaural Level Difference

The ILD operates in a frequency dependent manner. Feddersen et al showed that as the frequency of a sound increases (and the wavelength decreases) ILDs become greater (Feddersen et al. 1957). This is because when a sound's wavelength is shorter than the width of the head it no longer diffracts fully around the head. Feddersen's work also showed that ILDs may be as large as 20dB for high frequency sounds that are to the side of the listener.

The distance to the head can also affect the ILD. At small distances close to the head (less than about 1m) the wavefront curvature plays an important role. For example, Brungart and Rabinowitz show that sources very close to the head give greater interaural level differences than sources that are further away (Brungart & Rabinowitz 1999).

2.2.3 Interaural Time Difference and the Interaural Phase Difference

The ITD occurs because of the different path lengths (P_L and P_R) travelled by sound to both ears. It can vary from $0\mu s$ when a sound source is in front of the listener, to approximately $700\mu s$ when a sound source is at the side of the listener. The ITD is dependent on the distance d separating the listener's ears (see figure 2-2).

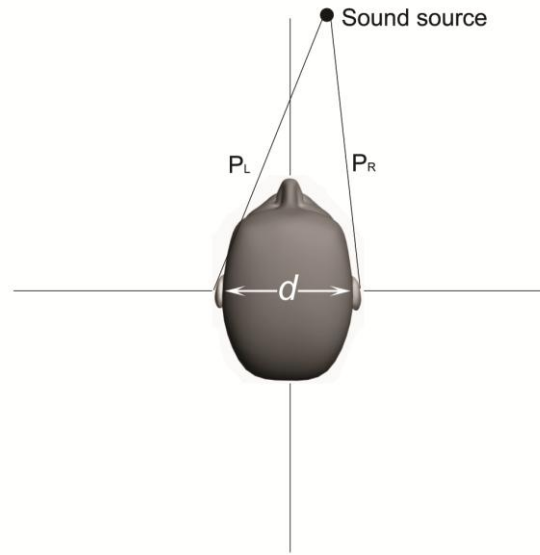


Figure 2-2: Interaural Time Difference

For a spherical head the ITD can be approximated at low frequencies using the following equation:

$$ITD = \frac{2r}{c} \sin \theta \quad (2.1)$$

where θ is the horizontal angle of the sound source, r is the radius of the head and c is the speed of sound (approximately 340 m/s).

The way in which the human auditory system decodes ITDs is dependent on the characteristics of the sound. For abrupt sounds the onset (or offset) differences between the signals at each ear are used (Blauert 2001). In this case the auditory system can extract useful information throughout the human hearing range. For continuous periodic sounds the auditory system decodes a time difference as a phase difference between the left and right ear signals (sometimes referred to as the Interaural Phase Difference (IPD) in the literature) (see figure 2-3). It should be noted, however, that IPDs are only usable up to about 1.5 kHz. Above this frequency, the wavelength is shorter than the diameter of the head rendering the phase information ambiguous. Moore points

out that this ambiguity lies in the fact the auditory system cannot detect absolute phase shifts (Brian Moore 2003).

Although the ITD and IPD are clearly related, they should not really be considered equivalent. A constant ITD leads to an IPD that varies linearly with increasing frequency. For example, an ITD of $500\mu\text{s}$ is equal to an IPD of 45° for a 250 Hz sine wave. For a 500 Hz sine wave, however, there is an IPD of 90° . Figure 2-3 illustrates this.

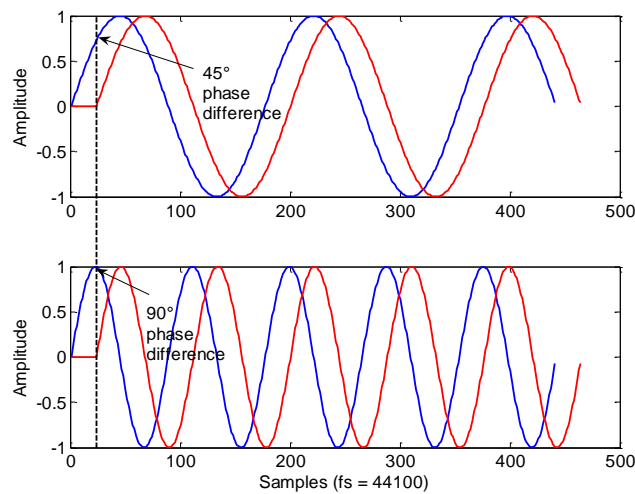


Figure 2-3: Interaural phase difference is dependent on frequency. Both waves in the top plot have a frequency of 250Hz whereas both waves in the bottom plot have a frequency of 500Hz.

Wightman and Kistler have shown that when a stimulus contains low frequency components the ITD cue is the dominant interaural cue (Wightman & Kistler 1992). In other words, the position of the auditory image can be determined by the ITD regardless of the ILD cue. Other more recent work which has placed the interaural cues in conflict has corroborated this finding (Ozcan et al. 2002; Ozcan et al. 2003; Jeppesen & Moller 2005).

2.2.4 Head Related Transfer Functions

The ITD and ILD are not enough on their own for localising sounds. For sound sources in the horizontal plane there are always two points around the listener with identical ITDs and ILDs.

For example, sound arriving from a source at 45° from the front in the horizontal plane will have an identical ITD and ILD as sound arriving from a source at 135° from the front. If the vertical plane is considered as well then there will be a whole series of points on the surface of a cone which have the same ITD and ILD (see figure 2-4). This is known as the cone of confusion (Mills 1972).

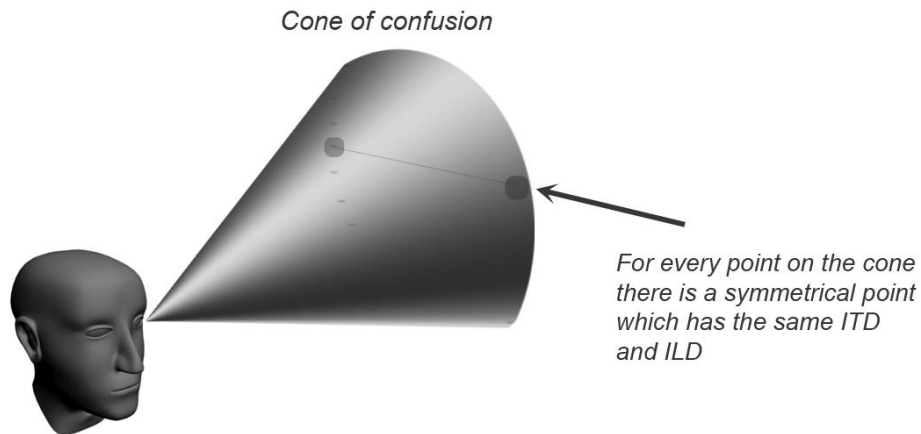


Figure 2-4: The cone of confusion

To resolve this ambiguity spectral cues are used which occur as a result of the directional-dependent filtering caused by sound reflecting off the ear's pinnae and upper body. A number of different studies have demonstrated that monaural (single ear) spectral cues are vital for the localisation of sound sources above and below the listener (Wright et al. 1974; Brian Moore et al. 1989). This has been clearly demonstrated in experiments by Gardner and Gardner (M. B. Gardner & R. S. Gardner 1973). There is also substantial evidence that the spectral cues incurred because of sound reflecting off the pinnae help us discriminate sounds coming from the front and back (Kistler & Wightman 1992; Blauert 2001; Langendijk & Bronkhorst 2002; Zahorik 2006).

Spectral cues can be described by a complex response function known as the Head Related Transfer Function (HRTF). A HRTF is defined as the acoustic transfer function measured between a sound source at a given location and the listener's eardrums. It is a frequency domain

function but has a corresponding time domain function known as the head-related impulse response (HRIR). The two domains can be related by the Fourier transform.

In engineering, HRTFs are usually specified as a minimum phase FIR filter with the ITD encoded into the filter's phase response and the ILD related to the filter's overall power (Cheng & G. H. Wakefield 2001). Typically, HRTFs are measured for both ears at different azimuths and elevations around the listener. The work by Han typifies this procedure (Han 1994).

Every individual has a personal HRTF for each sound source direction. The reason for this is because of our unique congenital features (i.e. head shape, head size, pinnae shape, pinnae size). To highlight this point, figure 2-5 plots the left and right ear HRTFs measured at azimuths 0° , 40° and 85° at an elevation of 0° for three different subjects. This figure demonstrates how much HRTFs can differ from person to person (especially at high frequencies) and also by angle.

Despite the uniqueness of HRTFs, however, common features can be found. For example, at mid/high frequencies sharp spectral notches occur because of sound interaction with the external and inner ear (in figure 2-5 this is most apparent around 8 kHz). There is strong evidence in the literature to support the hypothesis that these spectral notches are important cues for the localisation of sound in the vertical plane (B C Moore et al. 1989; Wright et al. 1974; M. B. Gardner & R. S. Gardner 1973; Bloom 1977).

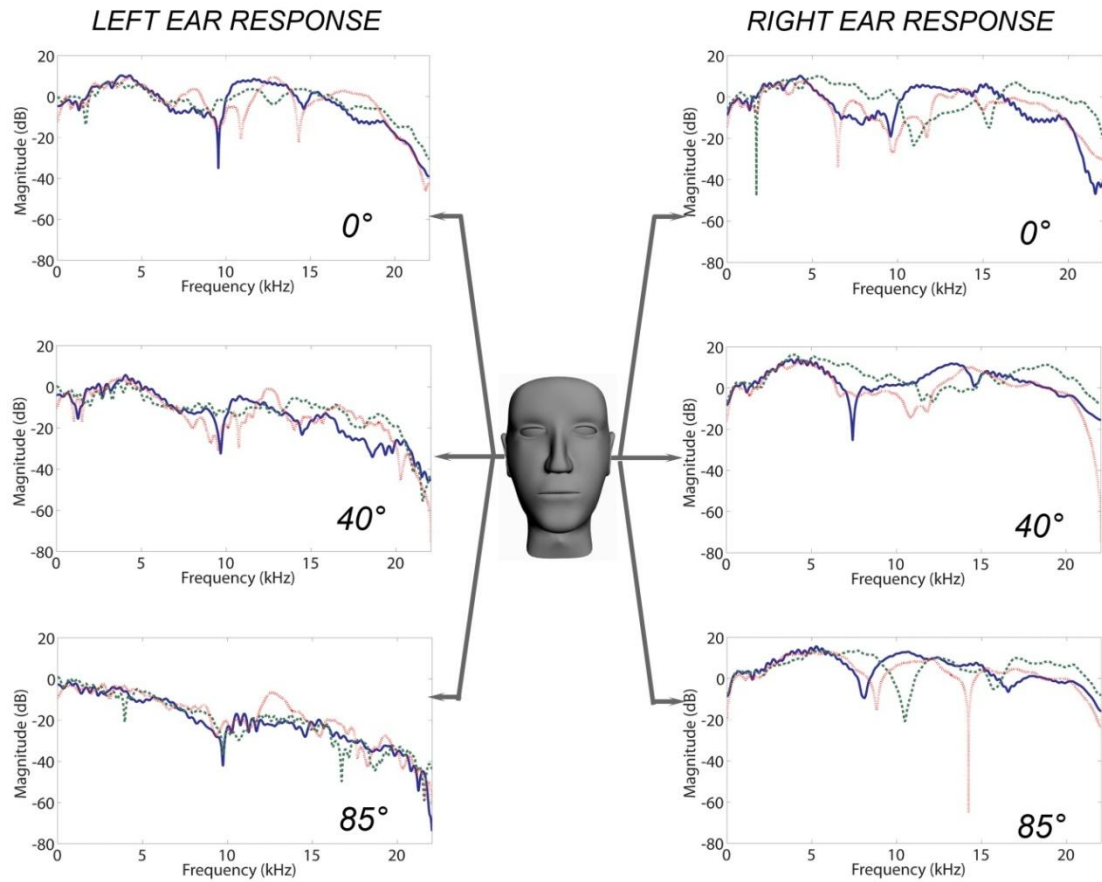


Figure 2-5: Left and right ear HRTFs for 3 subjects at 3 different angles (0° , 40° and 85°). These particular HRTFs were taken from the HRTF database generated by the Center for Image Processing and Integrated Computing (CIPIC 2004).

Since HRTFs are very individualistic, and the measurement of them is a time consuming matter, a significant amount of recent research in this area has explored ways of extracting and characterizing their common features (Katz 2001a; Katz 2001b; Zotkin et al. 2003; Dobrucki & Plaskota 2007). One of the main outcomes from this type of work is the creation of generic sets of HRTFs for application in Binaural synthesis (reproduction of 3D sound over headphones). Generic implementations, however, are under optimal and tend not to work very well in this respect because our auditory system is “tuned in” to our own HRTF (Wenzel et al. 1993).

One application of HRTFs particularly relevant to this work is in the extraction of auditory cues for assessing the localisation quality of sound reproduction systems. Using processing methods it is possible to derive auditory cues (i.e. ITD and ILD) directly from HRTFs to assess how capable a reproduction system is in reproducing auditory cues for a listener (Jot et al. 1995; Sontacchi et al. 2002; Nam et al. 2008). Jot et al demonstrated the effectiveness of this concept in the analysis of algorithms developed for surround sound reproduction (Jot et al. 1999); a similar method was also demonstrated to good effect by Wiggins when investigating the effects of listener head movement in sound localisation over Ambisonics sound reproduction systems (Wiggins et al. 2001).

2.2.5 Head movements

Localisation of sound sources is assisted further by head movements (Thurlow & Runge 1967; Noble & Gates 1985; Kato et al. 2003). Small head movements result in slight changes in ITD, ILD and spectral filtering, helping the listener focus in on the sound source. Head movements play a very important role when there is limited cue information available. For example, they can help us resolve front-back localisation confusion which can occur when limited localisation information is available from a sound (Wightman & Kistler 1999).

2.2.6 Localisation accuracy and spatial resolution

The accuracy with which we can localise sounds is of particular interest to this research. Systems developed as part of this work will be evaluated on their ability to correctly position sound sources around the listener. Numerous studies have shown that human localisation accuracy varies markedly with frequency (Stevens & Newman 1936; Blauert 2001; Brian Moore 2003). Generally, human localisation accuracy remains approximately constant for frequencies below 1 kHz. For frequencies between about 1 kHz and 3 kHz, however, acuity degrades somewhat until after 3 kHz when it improves again. The reason for degradation in this frequency region is because the interaural phase differences start to become ambiguous after 1 kHz, whereas below 3 kHz the interaural level differences are not always significant enough for a listener to lateralise a sound successfully. This problematic cross-over region can be seen in various studies (Blauert 2001; Brian Moore 2003).

Localisation has been shown to be most accurate directly in front of the listener (Blauert 2001). This accuracy decreases as the source moves to the side of the listener and improves again at the direct rear. The relationship between the angle of the sound source and accuracy of localisation is approximately the same for both low and high frequencies.

Human spatial resolution can be measured by asking a listener to determine the smallest noticeable difference in a change of a sound source's position. This difference is known as the minimum audible angle (MAA). Studies have shown that the resolution of the MAA is dependent on angle of the sound source around the listener (Mills 1958; Simon R Oldfield & Parker 1984; Saberi et al. 1991). A greater resolution is possible for frontal and rear sounds, with poorer resolution at the sides. Under ideal conditions (i.e. anechoic environment) a resolution of about 1° is possible for sounds at the front (Blauert 2001).

When reflecting on auditory localisation research and surround sound research it became clear that knowledge of human spatial resolution is not often considered during the development and analysis of surround sound systems. It appears that all systems in operation assume human hearing capability is equally capable in every direction. As psychoacoustic research has shown though, this is clearly not the case. This will be investigated further in chapter 4.

2.2.7 The precedence effect

The precedence effect says that listeners will tend to localise a sound source in the direction of the earliest arriving wave front. In one of the classic studies of the precedence effect, Wallach et al (Wallach et al. 1949) demonstrated how correlated sound waves arriving in close succession are fused together and heard as a single sound with a single location. Sound fusion is highly dependent on the nature of the sound. However, for short transient sounds it will occur within 1ms to 5ms, whereas for wide-band sounds, such as speech, it will occur within 5ms - 50ms (Litovsky & Colburn 1999). For sounds arriving before this time a single phantom source image will be perceived at a location determined by the contributing sounds (known as summing localisation). If sounds arrive after this time they will be heard as separate sources (i.e. echoes). Figure 2-6 illustrated these aspects of localisation.

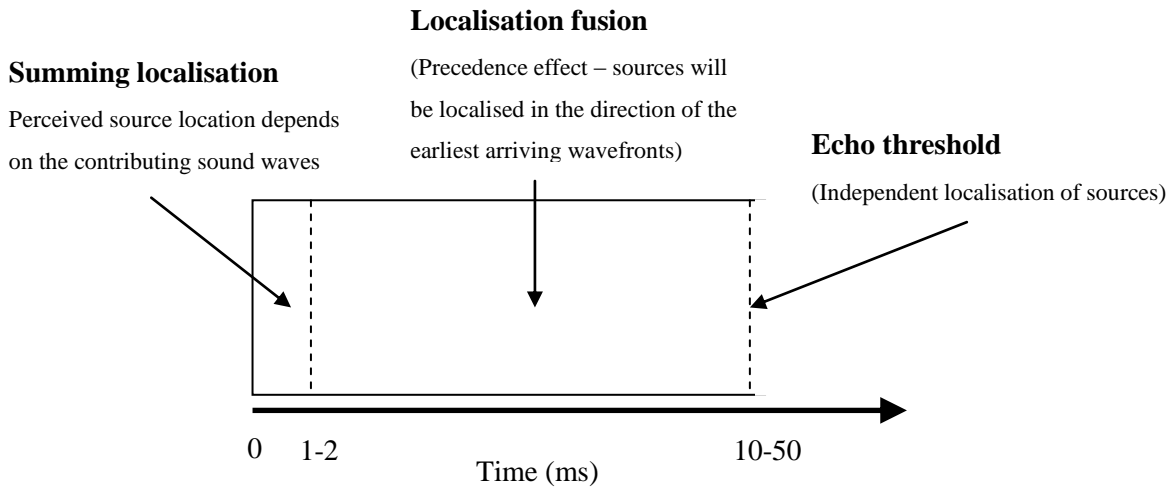


Figure 2-6: Summing localisation, localisation fusion and the echo threshold.

This effect is particularly important when localising sound in a reverberant environment due to the amount of information (reflections) reaching our ears simultaneously. It is also important when determining the location of sounds generated by multiple speakers in a surround sound system. Many surround sound techniques rely on sound emitted from the loudspeakers arriving synchronously at the listener's ears to generate the illusion of a phantom sound source. When sound does not arrive synchronously, however, the illusion can be lost due to the precedence effect taking over (see figure 2-7). Results from a recent surround sound localisation test by Bates et al demonstrate this (Bates, Kearney, Furlong et al. 2007).

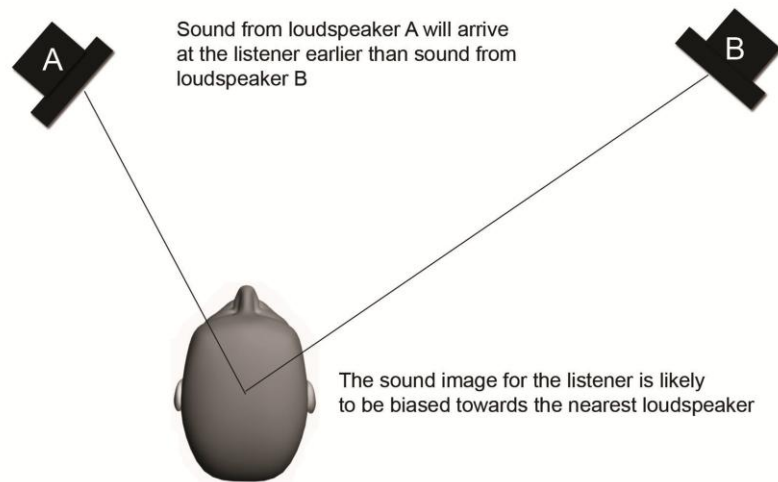


Figure 2-7: The potential problem of the precedence effect in a 2-channel listening situation

2.3 Surround sound

In the real world sound arrives from an infinite number of possible directions around the listener. In surround sound, however, there are typically only a finite number of loudspeakers in a finite number of directions. In order to match reality as closely as possible, surround sound systems attempt to replicate the auditory cues the listener would experience. A number of systems are capable of creating this illusion but before reviewing them and the methods used for analysing them, a historical perspective of surround sound will be given. Particular attention will be paid to how the industry has evolved to produce today's level of surround sound reproduction.

2.3.1 A historical perspective

Over the years the development of surround sound has mainly been driven by the film industry. The first significant multichannel sound system used with film was the "Fantasound" system developed for the 1940s film "Fantasia" by Walt Disney. Fantasound used a four channel optical soundtrack synchronised with the projected film. The soundtrack consisted of three audio channels, and a control track. The control track was used for distributing the audio to ten loudspeakers positioned around the audience. Despite the success of road-show presentations, Fantasound did not take off commercially because of the expense and logistics involved with implementing the system at the time (M. F. Davis 2003).

During the 1950s another elaborate multichannel system was developed for use with film. This system was known as Cinerama and was developed for the film “This is Cinerama”. It employed three synchronised projector screens, each covering one third of the total screen. The sound system that accompanied it used seven tracks stored on magnetic tape (six audio tracks and one control track). The loudspeaker system used for this film consisted of five frontal loudspeakers and an array of surround loudspeakers that could be fed a mixture of the source channels. Like Fantasound, this system was very advanced for the time, and consequently few cinemas used it because of the expense involved. Furthermore, there were few films being made at the time that would make full use of its capabilities.

In 1975 the surround sound industry was reinvigorated when Dolby Laboratories introduced “Dolby Stereo”. Dolby Stereo allowed the reproduction of 4 channels of audio from just 2 channels of data representing the left and right stereo signals (Dolby 1999). The 4 channels include a centre track, and left and right tracks for good frontal imaging and a mono surround track used for delivering ambience out of a number of loudspeakers distributed around the audience. Thanks in part to the success of blockbuster films such as Star Wars and Close Encounters of the Third Kind, the system was adopted for use in cinemas across the world.

Recognizing the potential for this technology in the domestic environment, in 1982 Dolby released a version of Dolby Stereo marketed as ‘Dolby Surround’. This was the first technology to be licensed to consumer electronics manufacturers as a means of decoding surround sound (Dolby 1999).

Not long after the success of Dolby stereo and Dolby Surround (during the 1980s) a channel configuration was agreed by the film industry that offered listeners an even more enhanced movie experience. The configuration includes a total of 6 channels: 5 full bandwidth channels for front left and right, centre and left and right surrounds, plus an optional low bandwidth channel for low frequency effects. This system is commonly referred to as 5.1 surround sound and is the standard channel configuration for mass market surround sound today. Although 5.1 has its origin firmly

placed in the film industry, it has been adopted for use in the music industry (Holman 2000), the video games industry (Ibbotson 2007), and radio broadcasting (Ternstrom 2003; AES 2004).

2.3.2 Industry standard surround sound loudspeaker configurations

In the early 1990s, the European standards organisation known as the ITU (International Telecommunications Union) began conducting research to determine the optimum speaker placement for a 5.1 system. This culminated in a document published in 1992 entitled "Recommendation for Multichannel Stereophonic Sound System With and Without Accompanying Picture" (ITU 1994), which details the now accepted industry standard surround sound loudspeaker configuration (figure 2-8).

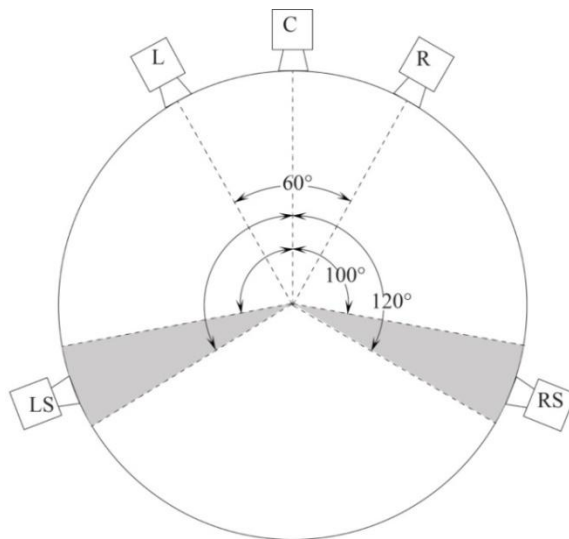


Figure 2-8: The standard ITU 5.1 loudspeaker arrangement.

The centre loudspeaker is placed straight ahead at 0° from the principal listening position. This loudspeaker was intended for pinning dialogue to the screen in movies (this is particularly useful for listeners in off-centre listening positions). The left and right speakers are located at $\pm 30^\circ$ in order to keep compatibility with existing stereo recordings and the surround loudspeakers are recommended to be placed between 100° and 120° . The decision for recommending the surround loudspeaker angles was determined from the results of experiments into the reproduction of sound images versus producing effect of envelopment for the listener (Gunther Theile 1991;

Gunther Theile 1993). Research has shown that decorrelated sound waves arriving at the listeners ears from the sides contribute significantly to the sensation of envelopment (Barron & Marshall 1981; Griesinger 1999; Blauert 2001).

Recent research by Muraoka and Nakazato on different 5-speaker configurations also supports the ITU 5.1 recommendation in terms of reproducing the soundfield at the ears of a centrally seated listener (Muraoka & Nakazato 2007).

Although the ITU standard clearly explains the optimum positioning of loudspeakers, it does not define anything about the way sound signals are represented or coded for surround sound. There is, in fact, no standard algorithm used for determining 5.1 loudspeaker feeds, rather a multitude of different algorithms.

Since the introduction of 5.1, other loudspeaker configurations have emerged which take advantage of increased bandwidth on next generation media formats (e.g. Blu-ray). These configurations include 6.1, 7.1, 10.2, and more recently 22.2 proposed by Hamasaki et al (Hamasaki et al. 2004). At the time of writing, 6.1 and 7.1 systems are mainly used as desktop computer surround sound systems, whereas the 10.2 and 22.2 tend to be used in large-scale listening situations such as in the cinema because of the quantity of loudspeakers used.

2.3.3 Typical surround sound loudspeaker configurations

Whilst conducting research into the use of 5.1 systems it became clear that few people followed the ITU guidelines when setting up a surround sound loudspeaker arrangement in a domestic environment. It appears that the loudspeakers are arranged in a manner which is convenient for the listener(s). Figure 2-9 and figure 2-10 show two different arrangements which might typically be used in a domestic environment.

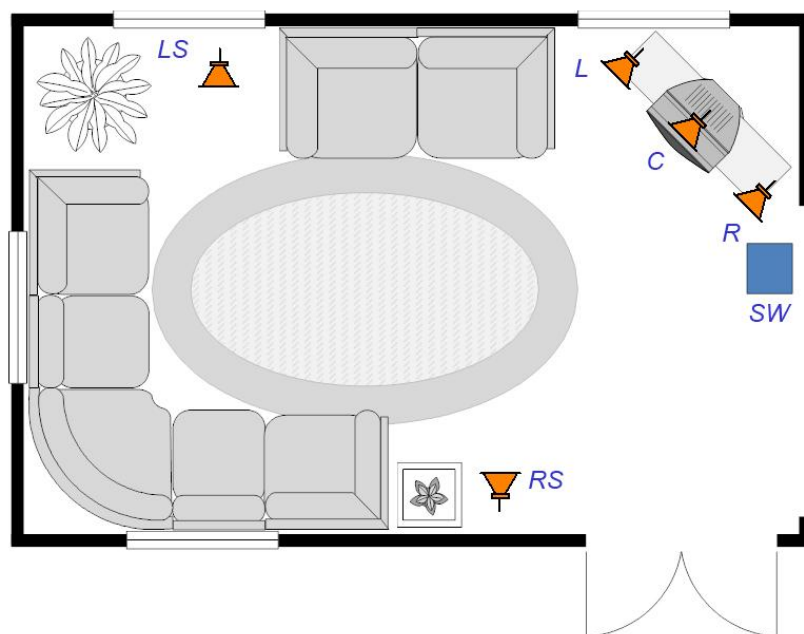


Figure 2-9: First example of a typical 5.1 setup in a domestic environment. L (left), R (right), C (centre), LS (left surround), RS (right surround), SW (sub woofer)

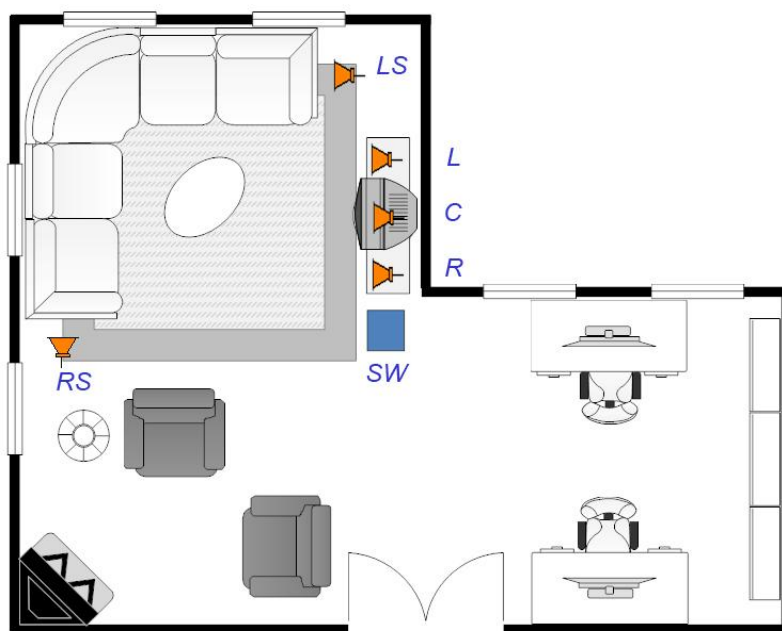


Figure 2-10: Second example of a typical 5.1 setup in a domestic environment. L (left), R (right), C (centre), LS (left surround), RS (right surround), SW (sub woofer)

Probably the main issue in setting up a 5.1 system according to the ITU standard is the placement of the rear loudspeakers. In a domestic environment, walls or furniture usually prevent the user from placing the rear loudspeakers in the correct positions (Gunther Theile 1993). As a solution, users typically opt to fit them in convenient positions around the furniture. This problem is apparent in both of the examples given above.

Considering this, it may be concluded that a technique for reproducing surround sound in a domestic environment must be robust enough to cope with irregular loudspeaker placement since the placement of loudspeakers according to standards is generally not user friendly unless the user has a dedicated space for setting up the surround sound system.

2.3.4 General review of surround sound reproduction techniques

There are a number of different techniques for the reproduction of surround sound over loudspeaker configurations. Each of which can be categorised into one of the following three areas:

1. Positioning of sound sources using inter-channel differences
2. Reproduction that takes into account the listener's congenital features
3. Wavefield reconstruction methods

There follows an overview of several different techniques.

2.3.4.1 Positioning of sound sources using inter-channel differences

A sound can be made to appear to come from between a pair of loudspeakers by outputting the sound from both of the loudspeakers. This is an auditory illusion that is often referred to as a phantom image. The position of a phantom image can be controlled by changing the ratio of amplitude differences or time differences between the loudspeaker outputs referred to as panning.

Amplitude panning involves using inter-channel sound level differences to position the phantom image between the loudspeakers (typically between 0dB and 30dB). This technique is perceived by the listener in a frequency dependent manner. At mid to high frequencies, where interaural

phase differences cannot be used by the auditory system, interaural amplitude differences caused by the shadowing effect of the head are used. This is, in fact, the underlying principle behind Blumlein's stereophonic system invention (Blumlein 1937). At low frequencies an amplitude difference between the loudspeakers results in ear signals which have the same overall level but different phase. Figure 2-11 illustrates this for a sound source panned in front of the listener.

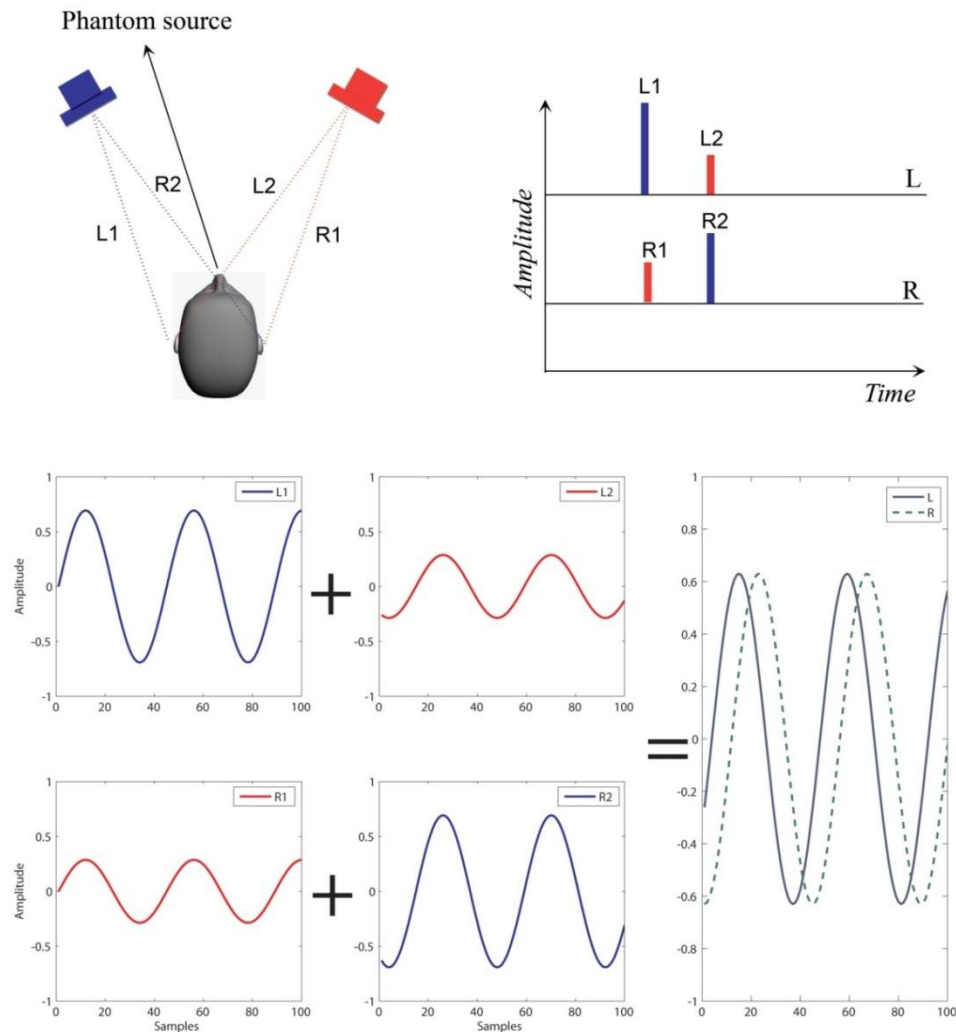


Figure 2-11: Inter-channel amplitude differences result in phase differences between the signals at the ears. The diagram is colour coded. Blue symbolises the left loudspeaker signal, whereas red symbolises the right loudspeaker signal.

One of the most common methods of amplitude panning is a cosine-sine law where a cosine and sine function are used for generating sound level weightings for a pair of loudspeakers:

$$\begin{aligned} \text{Left speaker} &= S \cos(\theta) \\ \text{Right speaker} &= S \sin(\theta) \\ 0 &\geq \theta \leq \pi/2 \end{aligned} \tag{2.2}$$

where S is the audio signal and θ is the angle in radians. This law has a constant sound power level when panning a source across the sound stage resulting in the listener perceiving the source at a constant distance.

An extension of the above cosine-sine law is the most commonly used method for 5-speaker layouts and is used almost exclusively in mixing desks and in software audio sequencers. Whilst it works reasonably well for positioning sound sources between closely spaced speakers, Theile and Plenge have shown that problems can occur with generating stable phantom images between loudspeakers angled further apart than about 60° (G. Theile & Plenge 1977). More recent work by Martin et al (Martin et al. 1999) found similar localisation issues at the sides and the rear of listeners during an experiment using the standard ITU 5-speaker configuration, as did Corey and Woszczyk (Corey & Woszczyk 2002). It appears that these issues can generally be attributed to conflicting auditory cues. Pulki and Karjallainen (Ville Pulkki & Karjallainen 2001) showed that the auditory cues generated by amplitude panning can indicate sources are in different positions. Benjamin and Brown (Benjamin & Brown 2007) have since shown this problem is significant in the mid-frequency range of human hearing. Clearly a more robust algorithm is needed for surround sound reproduction over existing standard surround sound loudspeaker layouts.

Phantom sound sources can also be positioned using time panning. In time panning small inter-channel time delays are used to position the phantom source between the loudspeakers (typically from 0ms to 1ms). For inter-channel time delays of about 1ms the sound will be perceived as

coming from the location of the loudspeaker radiating the earlier sound (i.e. the precedence effect). For greater time delays the image starts to become diffuse and spread out and can even be heard as two distinct sources (i.e. an echo). One of the main problems with time panning is that it suffers from unstable phantom imaging (as highlighted in experiments by Martin et al (Martin et al. 1999)). This is especially true for any listener in an off-centre listening position because of the different distances sound waves need to travel from each loudspeaker to reach the listener's ears. For this reason time panning is not considered a suitable technique for reproduction over multichannel systems when localisation is important.

2.3.4.2 Reproduction that takes into account the listener's congenital features

The Binaural technique was first introduced in the early part of the last century (Hammer & Snow 1932). By placing probe microphones at the entrance to each ear canal and recording onto a two channel medium, all the spatial cues (ILD, ITD and spectral) can be preserved. Consequently, when replaying the audio over headphones it is possible to perceive full three-dimensional surround sound. This effect is strongest when the recording is made with microphones placed in the listeners own ears (because of the individuality of HRTFs) (Moller et al. 1996) and when the listener's head is tracked to take account of head movements (Inanaga et al. 1995). Whilst the Binaural surround sound technique works well, binaural reproduction can only take place over headphones and is therefore not a suitable technique for this research.

Attempts have been made to reproduce binaural audio over a traditional stereo arrangement (i.e. two loudspeakers 60° apart). This technique is known as Transaural. In order for it to work correctly specialist algorithms need to be implemented that take into account the crosstalk of the loudspeakers. Binaural spatial cues are only preserved if left and right ear signals are kept separate. When audio is played out of the right speaker it is heard by both the right and left ear of the listener. Hence Transaural techniques need to cancel out the audio from the right speaker arriving at the left ear (and vice versa).

Cooper and Bauck have designed crosstalk cancellation algorithms for this system (Cooper & Bauck 1988). This process has been further refined by Kirkeby and Nelson (Kirkeby & Nelson

1997). However, good playback is only perceived over a very small area making this technique only suitable for a single listener. Even in the sweet spot the imaging tends to be quite fragile in the sense that small head movements can destroy the 3D illusion. Furthermore, due to the nature of crosstalk cancellation, it is currently difficult to extend this approach to multiple listeners simultaneously at different positions. Despite the drawbacks of Transaural it has been shown to be successful for 3D audio in desktop computing where the listener is usually stationary (Sæbø 1998).

Ambiophonics is a hybrid surround sound technique. It is similar in practice to the Transaural technique in that it still employs crosstalk cancellation filters. However, unlike Transaural, it is compatible with existing stereo, four-channel and even 5.1 recordings (Glasgal 2001). The basic principle of Ambiophonics is to provide the listener with as much psychoacoustically correct information as possible. It does this by positioning the main pair of loudspeakers directly in front of the listener angled apart by about 10° . These loudspeakers supply the listener with the direct sound and early reflections one would encounter in a real concert hall whilst at the same time limit colouration of signals arriving at the ears because of the limited cross talk. Additional speakers distributed around the listening area are used for immersing the listener in ambient sound. This system is capable of delivering 360° surround sound when an additional pair of loudspeakers is added to the rear. This is referred to as Panambiophonics. Although Ambiophonics is growing in popularity, it will not be considered in this research because there is currently no generic panning law for positioning sounds around the listener as in other systems. Moreover, there are no current methods for synthesizing material from scratch.

2.3.4.3 Wavefield reconstruction methods

Wavefield synthesis uses a horizontal array of closely spaced loudspeakers. It is one of the most accurate forms of surround sound playback as it allows the accurate reproduction of wave fronts in a space (Berkhout et al. 1992; Berkhout 1998). However, this technique is impractical in all but specialist installations because it requires a large number of loudspeakers (e.g. typically of the order of twenty loudspeakers or more). Furthermore, large amounts of computational processing power are needed to provide the loudspeakers with appropriate signals. For these reasons this technique will not be used in this research.

Another technique known as Ambisonics is built around perceptual models of localisation developed by Michael Gerzon (Gerzon 1974). The system is designed to take into account the fact that human hearing uses different mechanisms for sound localisation in different frequency ranges. This is one of the key advantages that Ambisonics holds over other techniques as it was designed with human perception in mind. Another advantage is the efficient hierarchical encoding scheme it employs. This scheme employs spherical harmonics for spatially sampling a soundfield. For example in a basic first order system (i.e. using first order spherical harmonics) only four channels of information are required for distribution and storage of a full-sphere soundfield, and only three for a horizontal soundfield (this is much fewer than other surround systems). Moreover, this encoding scheme is also easily expandable to allow more information about the soundfield to be stored in additional channels (Daniel & Moreau 2004). It has been shown that given enough channels it is possible to reconstruct a wavefield over a large area (Daniel et al. 2003).

Encoded Ambisonic soundfields can be manipulated in a variety of different ways. For instance it is possible rotate the whole soundfield about the X, Y and Z axes using rotational matrices (Malham 1987). It is also possible to zoom in on a soundfield by using a technique Gerzon termed dominance (Gerzon & Barton 1992; Chapman 2008). This flexibility would lend itself well in modern day surround sound application areas such as music, videogames and movies.

Ambisonics is not a new technology. However, in recent years there has been a growing interest in it because of its potential and flexibility within a wide number of application areas (Wiggins 2008). For example, Ambisonics employs an encoding/decoding model where it is possible to mix a 3D soundfield without *a priori* knowledge of the geometry of the loudspeaker array. This is an attractive feature especially when there is a growing demand for media to be shared between different application areas and different venues. Gaston highlighted its importance in a recent study that focused on the sharing of audio between planetariums (Gaston 2008).

Ambisonics was originally developed for playback over regular loudspeaker arrays (where loudspeakers are placed at the vertices of a regular polygon). The design of these systems is straightforward and well documented (Gerzon 1985; Benjamin et al. 2006). Unfortunately, the design of systems for irregular arrays like the standard ITU 5.1 arrangement is not so easy. A non-linear system of equations needs to be solved in order to produce a decoder that outputs suitable loudspeaker feeds (Gerzon & Barton 1992). Gerzon himself admitted that solving these equations mathematically was “tedious”. Recently, however, an alternative approach to mathematically solving these equations has been introduced (Wiggins et al. 2003; Craven 2003). Wiggins’ work involves using a heuristic search algorithm to optimise decoders at the sweet spot according to models of auditory localisation. This methodology is flexible in that it allows Ambisonic decoders to be developed for potentially any arrangement of loudspeakers and also according to any design criteria. A related approach was also investigated by Craven (Craven 2003).

Although good progress has been made by Wiggins in this area, there is scope for further developing this line of work. Specifically, there is potential for further improving localisation performance at the sweet spot and, perhaps more importantly, there is a need for a method for optimising localisation performance in non-central listening positions (as Wiggins highlights in the future work section of his PhD thesis) (Wiggins 2004).

2.3.5 Subjective comparisons between surround sound reproduction techniques

A number of studies have made subjective comparisons between various surround sound reproduction techniques. In one study by Guastavino et al, a subjective comparison was made between Ambisonics, Transaural and stereo (Guastavino et al. 2007). Eleven subjects took part in two different experiments. The first experiment investigated the spatial quality of the systems in terms of envelopment, immersion, representation, readability, and realism. The second experiment focused on the localisation quality. The results from these experiments showed that in terms of spatial quality Ambisonics performed well. Listeners rated the system the most “immersive” and “enveloping”. In terms of readability and localisation, however, the Ambisonic system did not perform as well as the other techniques in this experiment. One possible explanation for the poor performance of Ambisonics in the localisation test could be the type of

Ambisonic decoder implemented. The authors state they used an “in phase” decoder which is a decoder specifically designed for large scale playback whereas the rig used in the tests only had a radius of 2 metres. Moreover, in phase decoders are known to compromise localisation at the sweet spot for improved localisation in off-centre positions (Malham 1992). Since spatial quality and localisation were measured at the sweet spot, a more suitable Ambisonic decoder variant would have been more appropriate.

In another comparative test by Wiggins (Wiggins 2004) the following systems were evaluated for sound source localisation:

- First order Ambisonics over an 8-speaker regularly spaced layout
- Second order Ambisonics over an 8-speaker regularly spaced layout
- First order Ambisonics over the standard ITU 5-speaker layout
- Pair wise panning over the standard ITU 5-speaker layout
- Transaural using two speakers at $\pm 5^\circ$

The results for this test show that the second order Ambisonic system performed the best in terms of localisation with the other systems giving comparable performance. However, Wiggins states that the Ambisonic decoder used for the 5-speaker layout was not optimised, implying that better performance could be achieved.

Kearney et al recently compared the localisation performance of several surround sound techniques in a concert hall environment (including First Order and Second Order Ambisonics) (Kearney et al. 2007). Nine subjects were asked to localise reproduced sound sources at different angles from different off-centre listening positions. The results showed that all surround sound techniques suffer from sound images being biased towards the nearest loudspeaker in off-centre positions. However, these tests demonstrated that the second order Ambisonic decoder used in the test was able to reduce this effect to some extent when compared with a first order Ambisonic

decoder. It must be noted here that none of the systems under evaluation in this test were optimised for off-centre listeners.

Other recent subjective tests have focused solely on Ambisonics. Benjamin et al tested the “real world” localisation performance of several different first order decoders designed for use with regular loudspeaker arrays (Benjamin et al. 2006). Their study highlighted some interesting and relevant points with regard to the reproduction of recorded material for a centrally seated listener: in particular, how sound images are generally more stable when a narrower angle is used between the frontal loudspeakers. However, all tests were limited to a few subjects (the authors) and no quantitative data was presented. The paper drew conclusions from the individual experiences of the listeners.

Apart from the work discussed above by Wiggins there does not appear to be any literature detailing listening tests carried out on Ambisonic decoders optimised for irregular loudspeaker arrays. Clearly there is a need for work in this area as irregular loudspeakers arrays are the most commonly used domestically.

2.3.6 Objective measures for evaluating surround sound systems

Several objective measures have been developed which provide a means of predicting the spatial quality of sound reproduction systems. These measures are important when assessing surround sound systems in development, or when making comparisons between systems before conducting subjective tests which are time consuming.

2.3.6.1 Models of auditory localisation

A number of studies have developed mathematical models of auditory localisation to aid in the analysis of sound reproduction systems. These models provide a means of predicting the perceived direction of sound sources and so are especially important for theoretically determining the localisation error in a reproduction system.

Clark et al developed one of the first mathematical theories for quantifying a sound reproduction system's performance (H. A. M. Clark et al. 1958). They show that for a stationary head, situated at equal distance from a pair of loudspeakers, it is possible to derive a simple localisation law that can be used to predict the perceived direction of a low frequency reproduced sound source given the magnitude of the loudspeaker gains, the angle subtending the loudspeakers and the distance separating the listener's ears. The law is based on the fact that at low frequencies in stereophonic listening loudspeaker amplitude difference results in phase differences at the ears of the listener (as highlighted in section...). In their paper they used this method for evaluating the low frequency localisation performance of Blumlein's 2-channel stereophonic system.

Shortly afterwards, Clark, Dutton and Vanderlyn's work was expanded further by Bauer (Bauer 1961) into what is now commonly referred to as the "Stereophonic Law of Sines" (see equation 2.3).

$$\frac{\sin\theta_l}{\sin\theta_A} = \frac{(S_l - S_r)}{(S_l + S_r)} \quad (2.3)$$

where θ_l is the angle of the virtual sound source as perceived by the listener between the angle subtended by the loudspeakers, θ_A is the angle of the real source. S_l and S_r are the gains of the left and right loudspeakers respectively. This law shows that by applying appropriate positive loudspeaker gains the angle of a virtual sound source for a centrally seated listener can be moved anywhere between the loudspeakers.

More recent work by Bernfeld expanded this theory for use in multichannel systems (Bernfeld 1975). Bernfeld showed that for symmetrical loudspeaker layouts with each loudspeaker equidistant with respect to the listener the following law can be used:

$$\sin \theta = \frac{A \sin \theta_A + B \sin \theta_B \dots + N \sin \theta_N}{A + B + \dots + N} \quad (2.4)$$

where θ is the perceived angle of the virtual sound source, A, B, \dots, N are the gains of the loudspeakers at angles $\theta_A, \theta_B, \dots, \theta_N$. So for a 4-speaker square arrangement of loudspeakers the perceived angle of the sound source could be calculated thus:

$$\sin \theta = \frac{\sqrt{2}}{2} \frac{LF + LB - RB - RF}{LF + LB + RB + RF} \quad (2.5)$$

with LF, LB, RB and RF representing the gains of the left-front, left-back, right-back and right-front loudspeakers respectively. Various subjective tests have demonstrated that the law of sines correlates well with real sound source localisation and is able to predict the perceived location of low frequency reproduced sound sources with a reasonable degree of accuracy (Leakey 1959; Benjamin 2006).

Similar low frequency localisation models have been developed. Makita (Makita 1962), Bernfeld (Bernfeld 1975), and Cooper and Shiga (Cooper & Shiga 1972) developed a method which takes into account the movement of the listener's head. Makita's work in particular, demonstrates that at low frequencies the perceived direction of the sound is in the direction of the velocity it produces. His model assumes that the location of the sound source is the angle the head must face in order for there to be zero interaural phase difference at low frequencies.

In all of the above models the listener's head is approximated by two spaced ears with no acoustic shadow from the head. Furthermore, the complex behaviour of the soundfield near to the head is not considered. Therefore, these models are only valid at low frequencies for the ITD and IPD cues.

For mid to high frequencies, where head shadowing causes ILDs, a different approach must be used. It involves examining the directional behaviour of the energy field in the area around the listener's head. De Boer describes one such model (Boar 1940). Another model which can be used for predicting mid-high frequency localisation is described by Damaske and Ando (Damaske & Ando 1972). Their model employs the use of the cross-correlation function to determine the degree of coherence between the left and right ear signals of a dummy head placed at the listening position. This has been termed the Interaural Cross-correlation (IACC). Highly correlated signals indicate sharp directional perception, whereas low interaural coherence indicates the sound image will be diffuse and hard for the listener to pin-point. It is possible to derive the perceived position of the sound source from the IACC by finding the maximum point in the output of the IACC. Various studies have since shown that this method can be calculated across frequency bands (ISO 1997; Muraoka & Nakazato 2007). This is termed the Frequency-Dependent Interaural Cross-Correlation (FIACC).

In a metatheory of auditory localisation Michael Gerzon describes a hierarchy of models that can predict the location of sound sources in different frequency regions (Gerzon 1992a). The two simplest, and possibly most important models described are the acoustic particle velocity model, which corresponds to Makita's low frequency localisation model, and the acoustic energy-flow model, which corresponds to De Boer's mid/high frequency localisation model. Gerzon points out that practically all models of auditory localisation (including the other elementary models described above) are special cases of these two models.

In his metatheory Gerzon derived a "localisation" vector for the velocity and energy models that can be used when designing sound reproduction systems. The angle of each vector is used to show the perceived direction of a reproduced sound source and the magnitude is an indicator to the quality of the reproduced sound image. A nominal value of one for the magnitude of both vectors is equivalent to a real single point sound source, less or more than this can be interpreted as a lack of precision in sound localisation by the listener. If both vectors are the same for a reproduced sound source as they are for the real sound source then the reproduced sound source should be perceived to be the same as a real sound source. These vectors have been used in many

studies to evaluate the performance of multichannel systems, see for example (Gerzon & Barton 1992; Daniel et al. 1998; Pernaux et al. 1998; Jot et al. 1999; Wiggins et al. 2003; Craven 2003; Wiggins 2004; Wiggins 2007). Furthermore, they are the very principle behind the design of Gerzon's Ambisonic technique (Gerzon & Barton 1992).

Gerzon derived other more advanced criteria in his metatheory that can be used to predict sound timbre or sound colour at the listener's ears. However, there is no evidence in the literature that these models are actively applied in reproduction system design and analysis. This could be in part due to their complexity.

More recent research work has looked at creating Binaural models for evaluating surround sound systems. Pulkki et al developed a computational binaural model that incorporated the effect the external and inner ear have on sound (Ville Pulkki et al. 1999). The model was shown to be able to predict various sound localisation phenomena in loudspeaker listening at low and high frequencies. For example, the model predicted that the localisation error of virtual sound sources is greater for high frequency sounds. A later publication by Pulkki demonstrated the use of this model to good effect when evaluating various 2D and 3D surround sound reproduction techniques (Ville Pulkki 2001). Similar theoretical models have been created by Sontacchi et al (Sontacchi et al. 2002) and also Braasch (Braasch 2005).

Binaural evaluation of systems can also be undertaken practically using binaural microphones. Mac Cabe and Dermot tested the localisation ability of several surround sound reproduction techniques by recording a pseudorandom sequence of noise with a binaural microphone placed at the central listening point (Mac Cabe & Furlong 1994). From the recorded data they were able to derive the ITD and ILD for the binaural microphone. The ITD and ILDs derived for each system under test were compared with ITD and ILDs derived when recording a real sound source around the microphone. The advantage of using this method is that it allows the "real world" performance of a system to be investigated.

2.3.6.2 Soundfield reconstruction analysis

The measurement of a system's performance can be approached from a different viewpoint. It involves analysing how well as system is able to recreate an actual soundfield within an area.

In 1987, Vanderkooy and Lipshitz presented a paper where the performance of a stereo system and a first-order Ambisonic system were judged on their ability to recreate a theoretical 2D plane wave within the vicinity of the central listening point (Vanderkooy & Lipshitz 1987). The measure of error that they used is termed the integrated wavefront error and originates from the work of Bennett et al (Bennett et al. 1985). Basically, it involves integrating, over a circular path with radius r around the central listening area, the magnitude of the complex difference between the total acoustic pressure wave generated by the loudspeakers, and a theoretical plane wave travelling through the reproduction area i.e.

$$D(kr, \psi) = \frac{1}{2\pi|P_\psi|} \int_0^{2\pi} |S(kr, \phi) - S_\psi(kr, \phi)| d\phi \quad (2.6)$$

where $D(kr, \psi)$ is the integral wavefront error over a circular path with a radius r around the origin, P_ψ is the pressure of the reference plane wave, k is the wave number, $S(kr, \phi)$ is the pressure wave generated by the N loudspeakers and $S_\psi(kr, \phi)$ is the pressure wave generated by a best fit comparison plane wave. In a best case, the error $D(kr, \psi)$ would be zero for any length of r . However, Vanderkooy and Lipshitz's demonstrate that in practice the error in a system tends to increase with frequency and distance r from the centre point. In a further development of this work, Bamford and Vanderkooy analysed the integrated wavefront error for higher order Ambisonic systems (Bamford & Vanderkooy 1995). His work showed that by increasing the order of Ambisonics the wavefield can be reconstructed over a larger portion of the listening area.

While the above method is effective in determining a theoretical measure of a system's performance, it assumes for mathematical simplicity that the loudspeakers are situated in a free

field (an environment with no sound reflections). It also assumes for mathematical simplicity that the sound waves arriving at the listening area are all perfect plane waves. In reality, however, this will not be the case unless the loudspeakers are all at an infinite distance from the central listening point. Nevertheless, this approach provides a basis with which a surround sound system can be compared with other surround sound systems under ideal conditions.

A number of other soundfield analysis approaches have been defined. One method considers the synthesis of a soundfield by matching the spherical harmonic amplitudes of the desired field with the sum of the spherical harmonic amplitudes produced by an ideal set of loudspeakers (Vanderkooy & Lipshitz 1987; Ward & T.D. Abhayapala 2001). Betlehem and Abhayapala have also developed a method of analysing a 2D reproduced soundfield in a reverberant environment (Betlehem & Thushara D. Abhayapala 2005). This has more recently been expanded to 3D soundfields by Poletti (Poletti 2005). While all of these methods differ in principle, many have an equivalent mathematical background.

Although soundfield analysis methods are interesting, they do not involve psychoacoustics in any way. The school of thought is that if the soundfield can be correctly formed within an area, any listener situated in that area should receive the correct psychoacoustic cues.

2.4 Optimisation using computer search algorithms

Computers are commonly used to solve complex problems. One method involves using a computer search algorithm to seek out the best parameters for a given problem. In general, this methodology is used when finding a solution mathematically is too difficult.

The application of computer search algorithms is wide ranging. They have been used to solve problems from a large number of different disciplines including physics, chemistry, mathematics and engineering. As highlighted earlier, computer search algorithms have been used in audio engineering research for developing surround sound systems. Wiggins (Wiggins et al. 2003;

Wiggins 2004; Wiggins 2007) and also Craven (Craven 2003) have used search algorithms to seek good surround sound decoder parameters according to models of auditory localisation.

Optimisation using computer search algorithms involves composing a function to measure the “fitness” of a set of parameters generated by the search (i.e. a fitness function). The fitness function can contain a single objective or multiple objectives to represent the key elements of the optimisation problem. These two approaches are referred to as single-objective optimisation and multi-objective optimisation respectively.

The simplest form of optimisation problem involves searching for the best single parameter for a single objective i.e.

$$f(x) = f_1(x) \quad (2.7)$$

where f_1 is an objective and x is a parameter. In most real world problems, however, there are often multiple parameters and multiple objectives to describe the key criteria of a problem. When this is the case it involves combining the values of all objectives into a single scalar fitness function. This can be written mathematically like so:

$$f(x) = [f_1(x) + f_2(x) + \dots f_n(x)] \quad (2.8)$$

where n is the number of objective functions and x is a vector of parameters corresponding to the dimensions of the search space. In general, finding a solution for a multi-objective fitness functions is more complicated than finding a solution for a single objective fitness function. This is because an improvement in one objective can result in the decrease in performance of another (i.e. objectives can conflict).

In a multi-objective fitness function weightings can be applied to the objectives i.e.

$$f(x) = [w_1 f_1(x) + w_2 f_2(x) + \dots w_n f_n(x)] \quad (2.9)$$

where w_n is the weighting applied to the n th objective. Applying weightings to individual objectives can give the user more control over the type of solution produced by the search. For example, applying a large weight to an objective can increase its importance relative to other objectives. On the other hand, applying a small weight to an objective can decrease its importance relative to other objectives.

A search algorithm can look for the minimum value of a fitness function or a maximum value of a fitness function depending on how the problem is configured. Although both methods are equally valid, the former approach will be used in this work (i.e. finding the best solution will entail searching for the *minimum point* in the domain of the fitness function). In order to illustrate the minimisation of a fitness function using a search algorithm figure 2-12 plots the search space of a test fitness function. This particular test function has two parameters which are optimised according to a single objective i.e.

$$f(x_1, x_2) = - \sum_{i=1}^2 \sin x_i (\sin i x_i^2 / \pi)^{20} \quad (2.10)$$

Where x_i is the i th parameter constrained within the range $0 \leq x_i \leq \pi$.

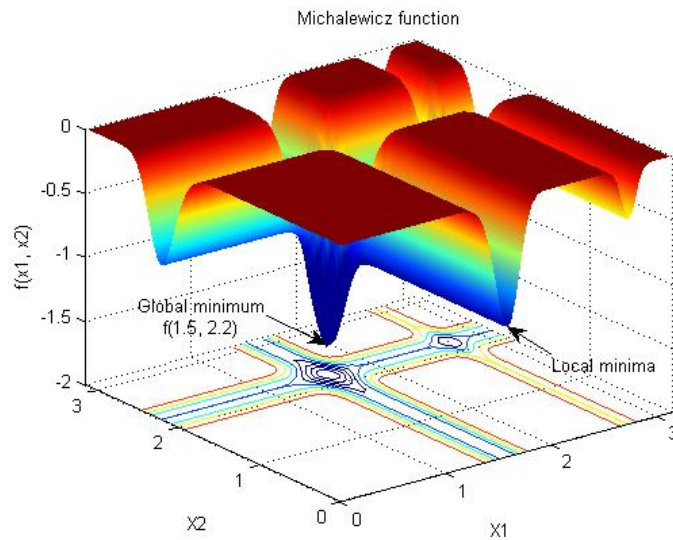


Figure 2-12: A 2D test function known as the Michalewicz function (Michalewicz 1998)

For this particular function it is clear that there is only one “optimum” solution (located at $x_1 = 1.5$ and $x_2 = 2.2$). This point is referred to as the global minimum because it constitutes the lowest point, and hence the “best” solution, in the search space.

There are also several local minima which are situated at the bottom of valleys in the search space. A solution is defined to be a local minimum when there are no other solutions within the vicinity with better fitness function values. This type of solution could be accepted by a search algorithm as a “good” solution.

By plotting functions in this way it is possible to visualise the search space and locate the region where the global minimum can be found. It is difficult, however, to visualise multi-dimensional search spaces (i.e. > 3). The next section will discuss some of the search algorithms which can be used to find local and global minima when this is the case.

2.4.1 Search algorithms used for optimisation

2.4.1.1 Exhaustive search

It is possible to use a brute force approach to search exhaustively to find the best solution to an optimisation problem. This approach consists of systematically evaluating all possible solutions in the search space according to the problem's statement. The advantage of doing this is that the best solution is always guaranteed to be found. However, the drawback is the potential time it takes to complete a search. The size of the search space is proportional to the number of potential solutions. For example, consider a problem with 4 parameters each with the range $[0, 1]$. If each parameter is checked at a resolution of 1 decimal place with a step size of 0.1 then there will be a total of 14,641 potential solutions. However, if a parameter resolution of 2 decimal places is used with a step size of 0.01 the increase of potential solutions is exponential (i.e. 104,060,401). Clearly exhaustively searching for the best solution is not always feasible. This is especially true when the fitness function is complicated and time consuming to compute.

2.4.1.2 Heuristic searches

When the search-space is too large to search exhaustively, a heuristic search algorithm can be used. The basic idea of a heuristic search is to navigate intelligently a subset of the search-space. By searching in this manner the algorithm is almost certain to find good local solution in a reasonable amount of time. However, there is no guarantee the global solution will be found.

There are many different heuristic search algorithms each with their own advantages and disadvantages. Some of the more commonly used algorithms are:

- Random solution search

This method consists of randomly evaluating solutions in the search-space until a solution is found which is acceptable. This is arguably the most simplistic approach to optimisation but has a number of advantages relative to other search algorithms. The advantages include ease of coding the software algorithm, and considerable increases in the number of solutions evaluated within a set time (i.e. less time is spent locating a good

solution and more time is spent evaluating more solutions). The disadvantages are the algorithm is not intelligent and consequently less reliable to reach a good solution within a set time.

- Random step search

A variation of the random solution search is the random step search. The basic idea of this algorithm is to randomly step from the current solution along each of the dimensions of the search space (i.e. generate a list of candidate nearby moves). Then choose the nearby solution with the best fitness score. This process is generally repeated until no improvement can be made.

- Steepest descent (also known as the gradient descent)

This algorithm is used to find the nearest local minimum in the search space. It starts at a random point in the search space and then moves in the direction with the steepest descent. In order for the algorithm to determine the direction of steepest descent the function must be differentiable. This method is guaranteed to converge on a local minimum. However, the local minimum it converges on might not be optimal. Furthermore, once a local minimum has been reached the algorithm becomes stuck with no method of escaping because it is always seeking a downhill gradient.

- Simulated Annealing

This search algorithm is modeled on the cooling process in annealing (a process of heating and cooling materials to change their properties) (Kirkpatrick et al. 1983). It is similar to the random step in that it replaces the current solution with a random nearby solution. However, the type of solution accepted depends on a tolerance parameter which is decreased during the search process. When the tolerance parameter is at a maximum then there is a high probability of accepting a worse move. As the tolerance is decreased the probability of selecting a worse solution decreases. This component of the search algorithm allows it to escape from local minima in the early stages of the search process.

- Tabu Search

The Tabu Search is as a meta-heuristic search algorithm. That is, it provides a framework for enhancing a local search by employing memory structures. For example, one of the memory structures (known as the Tabu List) holds a list of previous moves in the search space. This list is used to prevent the local search algorithm from revisiting areas that have already been explored. The advantage of using the Tabu List is the search can escape local minima (Glover 1989; Glover 1990).

- Genetic Algorithm

The Genetic Algorithm falls under a branch of heuristic searches based on the principles of natural evolution. During the search process the search maintains a population of possible solutions whilst trying to “evolve” better solutions by applying different processes modelled after evolutionary biology. These processes include inheritance, selection mutation and reproduction (Back & Schwefel 1993).

Other algorithms which are modelled on the natural phenomena include Particle Swarm Optimisation (after the social behaviour of flocking birds) and Ant Colony Optimisation (based on ants moving between their colony and a source of food).

Traditionally, heuristic search algorithms have been used to solve combinatorial puzzles (such as the classic N-Queens and the Travelling Salesman Problem). However, more recently there has been a large expansion of their use, with applications in Artificial Intelligence (Webster 1991) and planning (Biundo & Fox 1999), medical science (Westhead et al. 1997), dynamic programming and notably, in audio engineering by Wiggins (Wiggins et al. 2003).

Comparisons have been made between many search algorithms in order to gain knowledge about their effectiveness in solving problems. The algorithms are generally judged on the quality of solutions they produce and also on the computational effort they require to reach such solutions. Much of the data from this work, however, is inconclusive. For example, Rossi-doria et al found

that many of the more intelligent search algorithms (e.g. genetic algorithm, Tabu search) gave comparable performance, and that finding the best algorithm for a specific problem was difficult (Rossi-doria et al. 2002).

Despite this however, some algorithms seem to perform consistently well. The Tabu search is one such algorithm. It has been applied to many different problems with good results; a very small number of examples include (Misevicius 2005), (Battiti & Protasi 1997), (Dell'amico et al. 1999), and more recently (Gaspero & Schaerf 2007). The Tabu search dominates specific problems such as Job Shop Scheduling (JSS) (see Nowicki and Smutnicki for example (Nowicki & Smutnicki 1996)) and also Vehicle Routing problems (Gendreau et al. 1994). In a recent modern day application the Tabu search excelled in a GPS navigation problem (Saleh & Dare 2001). This particular algorithm is also tried and tested when developing surround sound decoders (Wiggins et al. 2003; Wiggins 2007).

2.5 High performance computing

High Performance Computing (HPC) is the use of computers to support scientists, engineers and other analysts in numerically intensive work, for example optimisation using computer search algorithms. It includes computing systems from workstations and servers to super-computers assigned to solve the some of the world's most demanding computational problems.

Currently, HPC implementation involves distributing a problem across multiple processors that operate in parallel. Breaking a problem down in this manner can result in significant increases in speed over traditional approaches where processes are run in series. One example of the power of HPC is in the prediction of complex weather patterns through advanced computer models (Wehner et al. 2008).

In the past, mainly due to expense, access to HPC systems has been restricted to large organisations and academic institutions. However, recently because of advances in computer processing power and accessibility HPC systems are becoming more readily available to the

general user. For example, a cluster of the latest video game consoles by Sony has recently replaced a supercomputer in one institute that seeks to solve problems in astrophysics (Khanna 2008). Companies such as Clearspeed have also developed products readily available to be used in conjunction with desktop computers (Clearspeed 2008).

Other methods of harnessing computer power are also becoming available through networking. One of these methods, known as “volunteer computing”, is fast becoming popular with home users. An open source project developed at Berkley University (BOINC) allows users to define their own problems and then invite people to share the computational load towards solving them. A recent paper by Anderson and Fedak demonstrates the potential of this paradigm (Anderson & Fedak 2006). This concept has also been implemented when developing a system capable of processing the data produced by the Large Hadron Collider (the world’s largest particle accelerator) at CERN, the European Centre for Nuclear Research. The data generated by the Large Hadron Collider is estimated to exceed 15 petabytes per year. To cope with this massive amount of information a system called the GRID captures and distributes the data for storage and processing at banks of computers around the globe (Segal et al. 2000).

Despite the growing availability of resources, HPC is not yet actively applied in Audio Engineering research. There is great potential for HPC applications in this field of research. For example, HPC could be used for the processing of complex soundfields in Wavefield Synthesis (Beckinger & Brix 2008), the complex modeling of acoustic spaces, and notably in this work, the development of surround sound decoders using computer search algorithms. Clearly, in the latter case HPC would lead to faster development of decoders and also potentially better solutions.

2.6 Summary

This chapter has examined four distinct areas of research relevant to the development of improved surround sound decoder algorithms. Firstly, psychoacoustic research was reviewed in order to highlight the different mechanisms our auditory system uses when deciphering sound information from our surroundings. It is clear that these mechanisms must be considered when developing a system for the reproduction of surround sound.

Section 2.3 reviewed the subject of surround sound. It was shown how the industry has evolved from early commercial applications of surround sound in cinema, to modern day applications ranging from personal music listening to computer games. The most common modern day systems in use are 5.1. Standard guidelines have been specified by the ITU for arranging the loudspeakers in a 5.1 system yet these guidelines are rarely followed in a domestic environment because of furniture or room constraints.

Pair-wise panning is the most commonly used technique for reproducing surround sound over the 5.1 loudspeaker array. Research has demonstrated, however, that this algorithm is sub optimal in some respects. Another method, Ambisonics, is a flexible, full system approach to surround sound. It benefits from being built around two well established models of auditory localisation and is also known to be capable of reconstructing a soundfield over a larger area than some other techniques. Ambisonics was originally designed for playback over regular arrangements of loudspeakers. However, relatively recent work by Wiggins has looked at using a computer search algorithm for deriving decoder coefficients that gives better performance over irregular loudspeaker arrays (such as the standard 5.1 arrangement). This previously unexplored avenue of research is important because for the first time it facilitated development of Ambisonic decoders for irregular loudspeaker arrangements.

Despite the early advances in this area, further exploration is needed to investigate the full capabilities of developing Ambisonic decoders using this method. Research is also needed to confirm the subjective performance of irregular Ambisonic decoders for centrally seated listeners and off-centre listeners who, in large-scale listening situations, will account for the majority of the audience.

Performance of surround sound systems can be assessed objectively in a number of ways (as was discussed in section 2.3.6). When assessing a system's ability to reproduce localisation cues for a centrally seated listener, the velocity and energy models can provide information about the perceived direction of sound sources. Despite the maturity of these models, however, they do not

currently take into account a system's ability to reproduce localisation cues for an off-centre listener.

Section 2.4 looked at optimisation using computer search algorithms. There are a number of search algorithms which can be used depending in the optimisation problem. When the search space is too large to search exhaustively for the best solution, a heuristic search algorithm can be used to locate good local solutions. Research has shown that many heuristic search algorithms have comparable performance. However, one algorithm in particular, the Tabu search, appears to perform consistently well in multi-objective optimisation problem. It is also tried and tested when developing surround sound decoders (see the work by Wiggins).

When implementing search algorithms on a computer, efficiency is the key to producing solutions quickly. One approach identified in the early stages of this research for improving performance is through the use of high performance computing. Using HPC in this context would reduce time-to-solution significantly which in turn would allow more solutions to be evaluated within a set time. This would ultimately increase the chances of finding a better quality of solution.

The remainder of this thesis will describe in detail a software-based design tool for producing improved Ambisonic decoders. The tool uses the Tabu Search algorithm for seeking decoder parameters that best fit psychoacoustic design criteria specified in a multi-objective fitness function. It also enables searches to be run locally on personal computers as well as remotely on HPC hardware for faster generation of solutions. The extensive evaluation of decoders produced by the tool will also be detailed.

Chapter 3

Background theory

3.1 Introduction

Before describing the decoder design tool and each of its components in detail, background theory will be provided on the techniques employed in this research. Section 3.2 will describe the velocity vector and energy vectors and will include their mathematical formulation. Section 3.3 will provide theory on first-order Ambisonic systems and higher-order Ambisonic systems (both of which are developed in this research). Finally, section 3.4 will give information about the Tabu Search algorithm used for seeking good Ambisonic decoder parameters according to developed fitness functions. At the end of each of these sections, a rationale will be given as to why each of these particular techniques was chosen.

3.2 Velocity and energy localisation vectors

The velocity vector and the energy vectors will be used in this research to quantify objectively the performance of developed Ambisonic decoders. As highlighted in the literature review, the vector magnitudes and angles can provide meaningful information about the perceived quality and direction of reproduced sound source image when give a system's loudspeaker gains and angles. The velocity and energy vectors in Cartesian form are formulated thus:

$$r_V^x = \sum_{i=1}^n S_i \cos(\theta_i) / P \quad (3.1)$$

$$r_V^y = \sum_{i=1}^n S_i \sin(\theta_i) / P \quad (3.2)$$

$$r_E^x = \sum_{i=1}^n S_i^2 \cos(\theta_i) / E \quad (3.3)$$

$$r_E^y = \sum_{i=1}^n S_i^2 \sin(\theta_i) / E \quad (3.4)$$

Where:

$$P = \sum_{i=1}^n S_i \quad (3.5)$$

$$E = \sum_{i=1}^n S_i^2 \quad (3.6)$$

r_V^x is the velocity vector in the x direction, r_V^y is the velocity vector in the y direction, r_E^x is the energy vector in the x direction, r_E^y is the energy vector in the y direction, n is the number of loudspeakers, θ_i is the angular position of the i^{th} loudspeaker and S_i represents the gain of the i^{th} loudspeaker. P is the pressure and E is the energy. Converting the vectors into polar coordinates yields their magnitude and angle i.e.

$$r_v = \sqrt{(r_V^x)^2 + (r_V^y)^2} \quad (3.7)$$

$$\theta_v = \tan^{-1} \frac{r_V^y}{r_V^x} \quad (3.8)$$

where r_v is the magnitude of the velocity vector and θ_v is the angle of the velocity vector.

$$r_E = \sqrt{(r_E^x)^2 + (r_E^y)^2} \quad (3.9)$$

$$\theta_v = \tan^{-1} \frac{r_V^y}{r_V^x} \quad (3.10)$$

where r_E is the magnitude of the energy vector and θ_E is the angle of the energy vector.

The velocity vector can be used for predicting a sound source's location and quality for audio frequencies below about 700Hz where interaural time differences and interaural phase differences are the dominant localisation cues. The energy vector can be used for predicting a sound source's location and quality for audio frequencies between about 700 Hz and 5000 Hz where the interaural level difference is the dominant cue (Gerzon 1992a). Note that if we express these two frequency ranges using a logarithmic scale then the velocity vector extends across more octaves within the human hearing range.

When measuring from a central listening position, the ideal angle for both vectors is when they match the intended angle of the reproduced sound source. An ideal magnitude for both vectors is unit magnitude. For an array of loudspeakers surrounding the listener, this level of magnitude is achievable for the velocity vector if sound is emitted from opposing loudspeakers with negative gains. However, for the energy vector, this level of magnitude is not possible. If two or more loudspeakers are fed with sound with non-zero gains then the energy vector magnitude will always be less the unit magnitude. This can be proved by observing the fact that the magnitude of the energy vector is an average of loudspeaker gains with *positive* values (note the square term in equations 3.3 and 3.4). Thus for the energy vector to have unit magnitude it would require each speaker to lie in the same direction as the sound source.

The velocity vector and energy vector are used in this research for the following reasons:

1. Both vectors correlate with the interaural time difference and the interaural level difference so will provide important information about how well developed surround sound systems perform in terms of providing psychoacoustic cues for the listener.
2. They provide a quick and efficient way of assessing candidate surround sound decoders produced by a search algorithm (efficiency in potential solution evaluation is key when using search algorithms).
3. They define the very nature of the Ambisonic system (shown in the next section).

3.3 Ambisonic theory

An Ambisonic surround sound system comprises an encoding stage and a decoding stage. This section will detail the theory behind both stages.

3.3.1 Encoding

In Ambisonics, soundfields can be encoded using a specially designed microphone or through direct multiplication with encoding functions. In the former case the Soundfield microphone is typically used. This microphone, invented by Gerzon and Craven (Craven & Gerzon 1977), employs four sub-cardioid capsules which are mounted on the face of a tetrahedron (see figure 3-1). The combination of these capsules enables sound to be captured in three dimensions of space.



Figure 3-1: The Soundfield microphone

The raw signals from the output of the microphone are known as A-format. After undergoing processing to compensate for the spacing between the capsules the signals are converted to a format known collectively as B-format which represents the captured soundfield:

$$\begin{aligned}
 W &= \frac{1}{2}(LF + LB + RF + RB) \\
 X &= \frac{1}{2}((LF - LB) + (RF - RB)) \\
 Y &= \frac{1}{2}((LF - RB) - (RF - LB)) \\
 Z &= \frac{1}{2}((LF - LB) + (RB - RF))
 \end{aligned}
 \tag{3.11}$$

where LF is the signal from the left front capsule, RF is the signal from the right front capsule, LB is the signal from the left back capsule, RB is the signal from the right back capsule and W , X , Y , and Z are the B-Format components (Rumsey 2001).

The B-format components correspond to zero and first order spherical harmonic functions. The zeroth order W component is a pressure signal equivalent to the output of an omni-directional microphone. The first order X and Y and Z components correspond to velocity microphones (figure of 8) parallel with the coordinate axes in 3D Euclidean space (figure 3-2 shows the response of W, X and Y components).

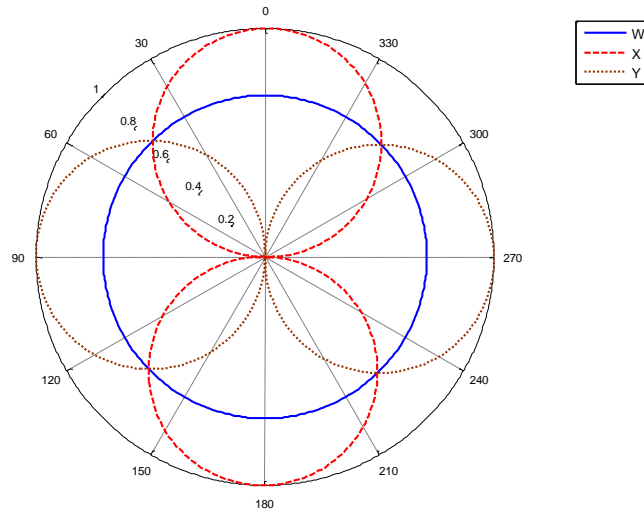


Figure 3-2: The angular response of the W, X and Y B-format components

In theory it is possible to capture a sound field using a higher order microphone (Cotterell 2002; T.D. Abhayapala & Ward 2002). However, at the time of writing this thesis, no higher order Ambisonic microphones are available commercially. This may change in the near future though as this is an active field of research (Moreau et al. 2006).

Ambisonic sound can also be encoded by direct multiplication with the encoding functions. To synthesize a first-order soundfield, for example, it is simply a matter of multiplying a monophonic signal with the following encoding functions:

$$\begin{aligned}
W &= S \frac{\sqrt{2}}{2} \\
X &= S \cos \theta \cos \varphi \\
Y &= S \sin \theta \cos \varphi \\
Z &= S \sin \varphi
\end{aligned} \tag{3.12}$$

with W , X , Y and Z , corresponding to the B-Format components, S the monophonic audio signal, θ is the azimuth of the sound source and φ is the elevation of the sound source. The weighting value of 0.707 is given for the W signal to allow for a more even distribution of levels within the channels (Craven & Gerzon 1977). In this work the focus is on improving surround sound reproduction in the horizontal plane so the Z component will not be used. Without the Z component W , X and Y can collectively be referred to as “horizontal B-Format”.

In order to expand the system to use higher orders, it is a simple matter of using the following equations for *horizontal* encoding:

$$\begin{aligned}
C_M &= S \cos M\theta \\
S_M &= S \sin M\theta
\end{aligned} \tag{3.13}$$

with C_M representing an additional component utilizing the cosine function, S_M representing an additional component utilizing the sine function, M is the system order and θ is the angle of the sound source in the horizontal plane. From this equation it can be seen that for every additional order the number of channels increases by two in a horizontal system. So for instance, a first order system employs three channels (W , X , and Y), a second order system uses five channels (W , X , Y , C_2 and S_2), a third order system uses seven channels (W , X , Y , C_2 , S_2 , C_3 , S_3) and so on. Figure 3-3 plots the horizontal encoding functions from first to fourth order. The encoding gains for each order are equivalent to the point where the sound source angle θ intersects the encoding function (i.e. the gain level is equivalent to the distance from the origin). It can be seen that by using higher order encoding functions there is a greater spatial resolution which leads to a

greater angular discrimination for sound sources when compared to lower orders (note θ_A and θ_B in each plot).

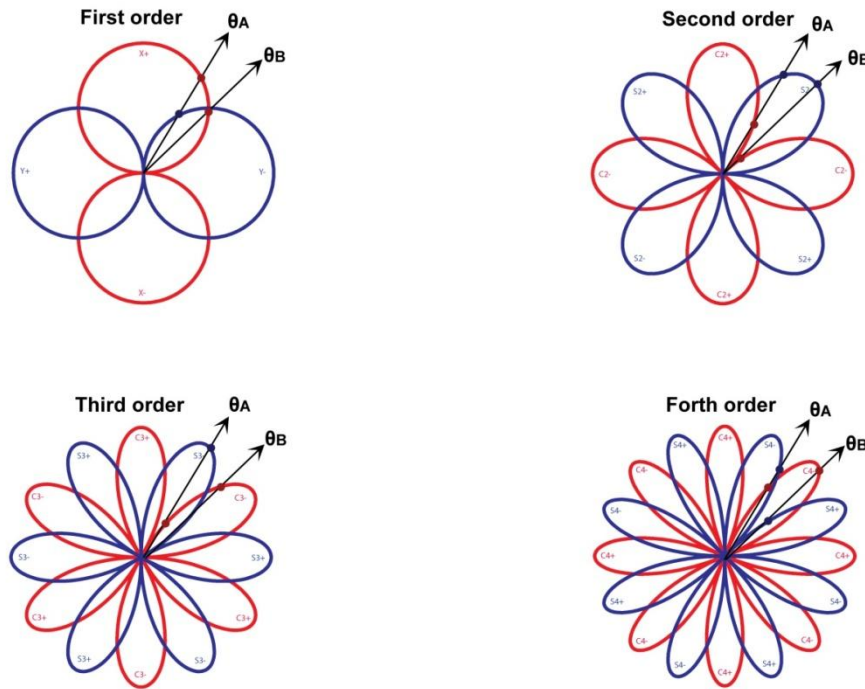


Figure 3-3: First order to forth order encoding functions

Once a soundfield is encoded ‘Ambisonically’ it is possible to manipulate it in a number of ways. For instance it is possible rotate or tilt the whole soundfield about the X, Y and Z axes using conversion matrices (Malham 1987). It is also possible to zoom in on first order soundfields by using Lorentz transformations (Gerzon & Barton 1992; Chapman 2008).

3.3.2 Decoding

Although Ambisonics is capable of reproducing a soundfield in three dimensions, this thesis focuses on sound reproduction in the horizontal plane. As a result, the scope of the theory presented here will be limited to decoding sound in horizontal plane.

3.3.2.1 Decoders for regular loudspeaker arrays

To decode ‘Ambisonically’ encoded audio a re-composition is made that takes into account the location of each loudspeaker. In order for the re-composition to constitute an Ambisonic decoding it must adhere to the following rules defined by Gerzon (Gerzon & Barton 1992):

- At the central listening position the velocity vector and energy vector angles match up until at least 4kHz
- At low frequencies (below about 400Hz) the magnitude of the velocity vector is ideal (unit magnitude)
- At mid/high frequencies (between 700Hz and 4kHz) the energy vector magnitude is substantially maximised across as large a part of the 360° sound stage as possible

For a decoder designed for a regular arrangement of loudspeakers (e.g. a square or a hexagon) it is straightforward to meet these requirements. To visualise why this is so, it is useful to use the concept of a virtual microphone. The virtual microphone is a simple way of understanding how encoding and decoding are related. Basically, each loudspeaker has a virtual microphone associated to it. The response of this virtual microphone is a weighted combination of the different encoding functions. The microphones point outwards from the central listening position as if they would directly capture the surrounding sound field. Their response describes the output level of each loudspeaker as a source is panned around the 360° sound stage.

For first order, the equation used to describe the response of each virtual microphone is given:

$$S_i = \frac{1}{2}[(2 - d)\sqrt{2}W + d(\cos \theta X + \sin \theta Y)] \quad (3.14)$$

where S_i is the speaker output, θ_i is the angle of the i th loudspeaker and d is the microphone directivity factor ranging from 0 to 2. A range of different first-order virtual microphone directivities is displayed in figure 3-4.

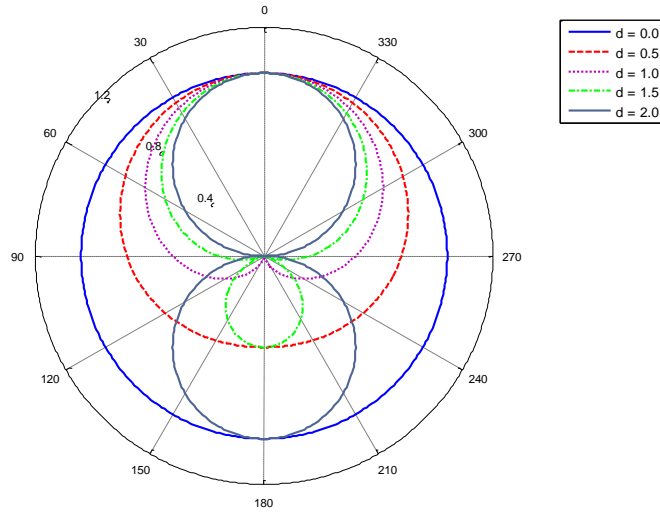


Figure 3-4: A range of first order virtual microphone directivities

By adjusting the directivity of each virtual microphone it is possible to optimise the velocity and energy vector responses for any regular loudspeaker array. There are three generic approaches to this depending on the type of velocity vector and energy vector response that is required: “basic decoding”, “max r_E decoding” and “cardioid decoding” (Moreau et al. 2006). Figure 3-5 displays the virtual microphone for each of these types of decoding for first order. Also displayed in figure 3-5 are the corresponding velocity and energy vector responses that would be obtained when using these decoders for a hexagonal arrangement of loudspeakers.

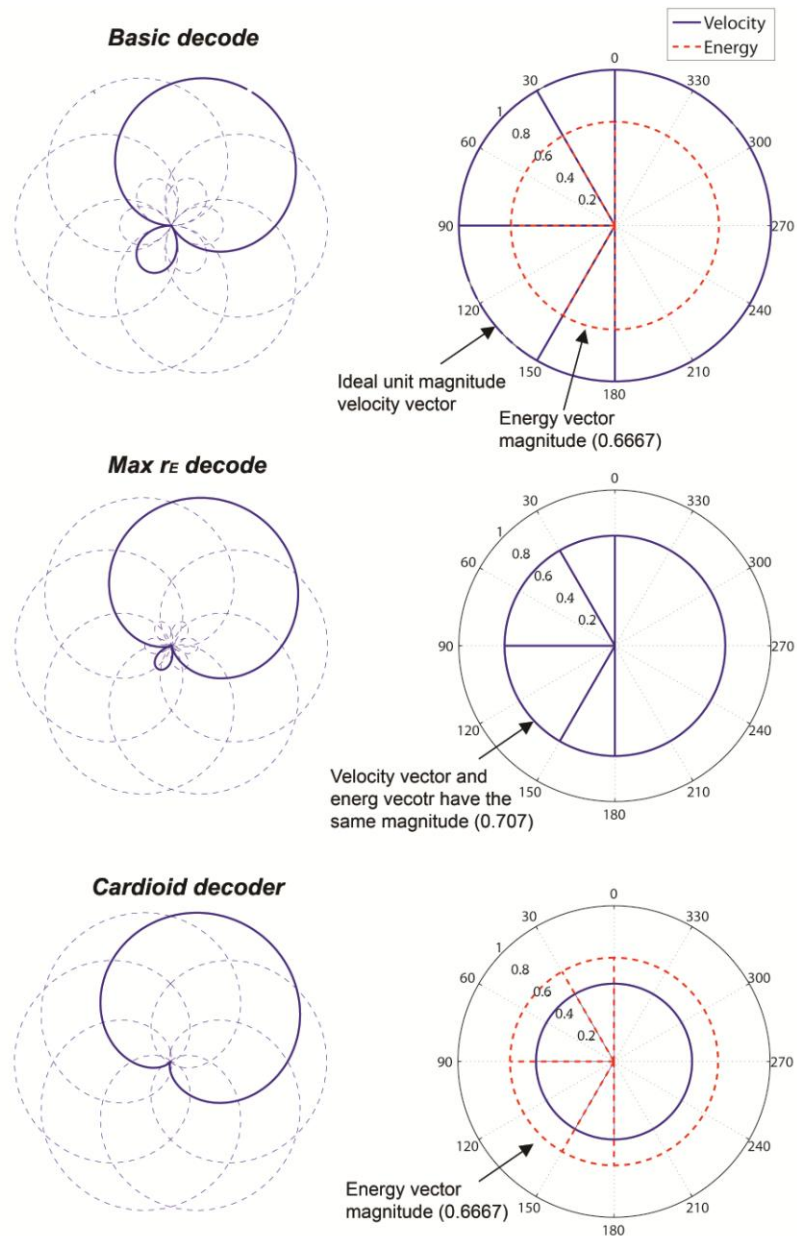


Figure 3-5: Three different types of Ambisonic decoding for a hexagonal loudspeaker array (Basic, Max r_E and Cardioid). The left column plots the virtual microphones for each type of decode and the right column plots the corresponding velocity vector and energy vector responses. In this plot the velocity vector and energy vector angles are shown at 0, 30, 90, 150 and 180 degrees (note they are ideal because they match the intended angles).

A basic decoding consists of maximizing the velocity vector response around the listener. In theory, if the Ambisonic soundfield produced by a basic decoder was recorded at the central listening position, it would match the originally encoded soundfield (Daniel 2001). A max r_E decoding consists of maximizing the energy vector performance. It does this by focusing the soundfield's energy in the expected direction (note the reduced size of the virtual microphone secondary lobes when compared with the basic decoding). Finally, a Cardioid decoder is specifically designed for large-scale listening (Malham 1992). For this type of decoder the virtual microphone secondary lobes are completely removed in order to limit the sound from the opposite loudspeakers to the sound source. This reduces the likelihood of listeners in off-centre positions localising reproduced sounds sources in the direction of the nearest loudspeaker. However, as a consequence of this, localisation performance at the central listening point is compromised (note the poorer velocity vector response in figure 3-5).

3.3.2.2 Decoders for irregular loudspeaker arrays

When developing an Ambisonic decoder for an irregular array of loudspeakers matters are not so straightforward. For example, if equation 3.14 is employed when designing a decoder for the ITU array then the performance becomes sub-optimal and does not meet Gerzon's original requirements for the Ambisonics system. To illustrate this, figure 3-6 plots the velocity vector and energy vector response for a first-order 'cardioid' decoder for this system.

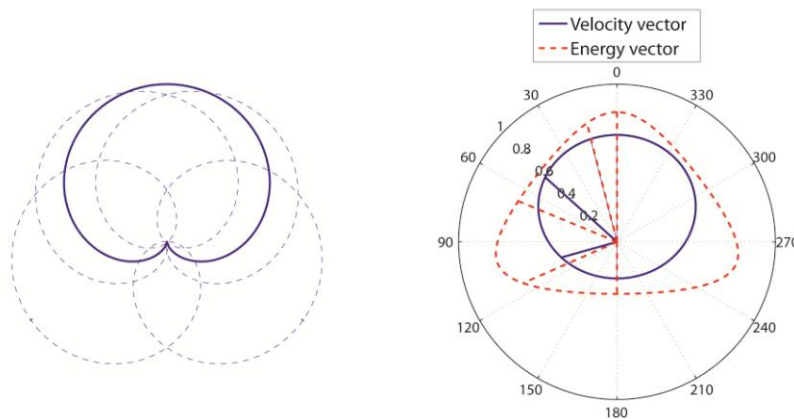


Figure 3-6: The performance of a 'cardioid' decoder for the ITU 5-speaker array. The localisation vector angles are shown at 0, 30, 90, 150 and 180 degrees (note they are not ideal).

As can be seen, angular distortion of the velocity vector and energy vector angles has been introduced and the magnitudes now vary by angle around the central listening point. Furthermore, using the same virtual microphone response for each loudspeaker results in a gain imbalance as a sound source is panned around the 360 degree sound stage. Sounds to the front will be louder than sounds to the rear because there are a greater number of loudspeakers in the front. Each of these anomalies is significant in terms of meeting Gerzon's requirements for the Ambisonics system. In terms of perceptual error Gerzon states that angle mismatch between the velocity vector and energy vector can reduce the focus of any reproduced sound source (Gerzon & Barton 1992).

In order to improve the velocity vector and energy vector response, and to correct for the gain imbalance, a different approach to decoding needs to be used that takes into account the irregular positioning of the loudspeakers. It involves using different weightings for each of the encoded components for each loudspeaker. The following system of equations describes this approach for an irregular left/right symmetrical 5-speaker first order decoder:

$$\begin{aligned}
S_1 &= (k_c^w \times W) + (k_c^x \times X) \\
S_2 &= (k_F^w \times W) + (k_F^x \times X) + (k_F^y \times Y) \\
S_3 &= (k_B^w \times W) - (k_B^x \times X) + (k_B^y \times Y) \\
S_4 &= (k_B^w \times W) - (k_B^x \times X) - (k_B^y \times Y) \\
S_5 &= (k_F^w \times W) + (k_F^x \times X) - (k_F^y \times Y)
\end{aligned} \tag{3.15}$$

where S_1 to S_5 are the gains of the centre, left, left surround, right surround and right speakers, k denotes a decoder coefficient, c , F and B denotes centre, front and back loudspeakers respectively, W , X and Y represent the horizontal B-format components of the soundfield. The values of the above coefficients are usually constrained within the range of 0 to 1 (Gerzon & Barton 1992). It can be seen that for this particular arrangement of loudspeakers 8 individual coefficients are

required. Note that if the left/right symmetry of the ITU array is broken 14 individual coefficients are needed.

Gerzon and Barton were the first to tackle the problem of deriving the above decoder coefficients for irregular loudspeaker arrays (Gerzon & Barton 1992). Their approach was to solve mathematically a non-linear system of decoding equations in order to find a suitable set of first-order decoder coefficients. However, Gerzon himself admitted that this was tedious and complicated (because of the square term in the energy vector equations). As highlighted earlier, an alternative method is to formulate the design of decoders as a search problem. This is the approach used in this research.

3.3.2.3 Additional decoding considerations

In Ambisonics it is possible to implement a dual-band decoding where the performance of the velocity vector and energy vector are optimised separately. The standard approach for a regular first-order decoder is to use linear phase shelf filters to adjust the level of the W signal in relation to the X and Y signals. These adjustments are made in the frequency regions where the velocity vector and energy vector operate (i.e. above and below approximately 700 Hz). By doing this one can take a system optimised for a basic decode and apply appropriate shelf filters to yield a system that has a max r_E decode at mid/high frequencies or vice versa (Lee 2005; Benjamin et al. 2006).

For an irregular decoder the concept is the same but the implementation is different. Rather than using shelf filters, a network of linear phase band-splitting filters is used with a different set of decoder coefficients in each frequency region (see figure).

In the literature dual-band decoders are often referred to as frequency-dependent decoders, whereas the single-band decoders are called frequency-independent decoders. This terminology will be used when describing such decoders in this thesis.

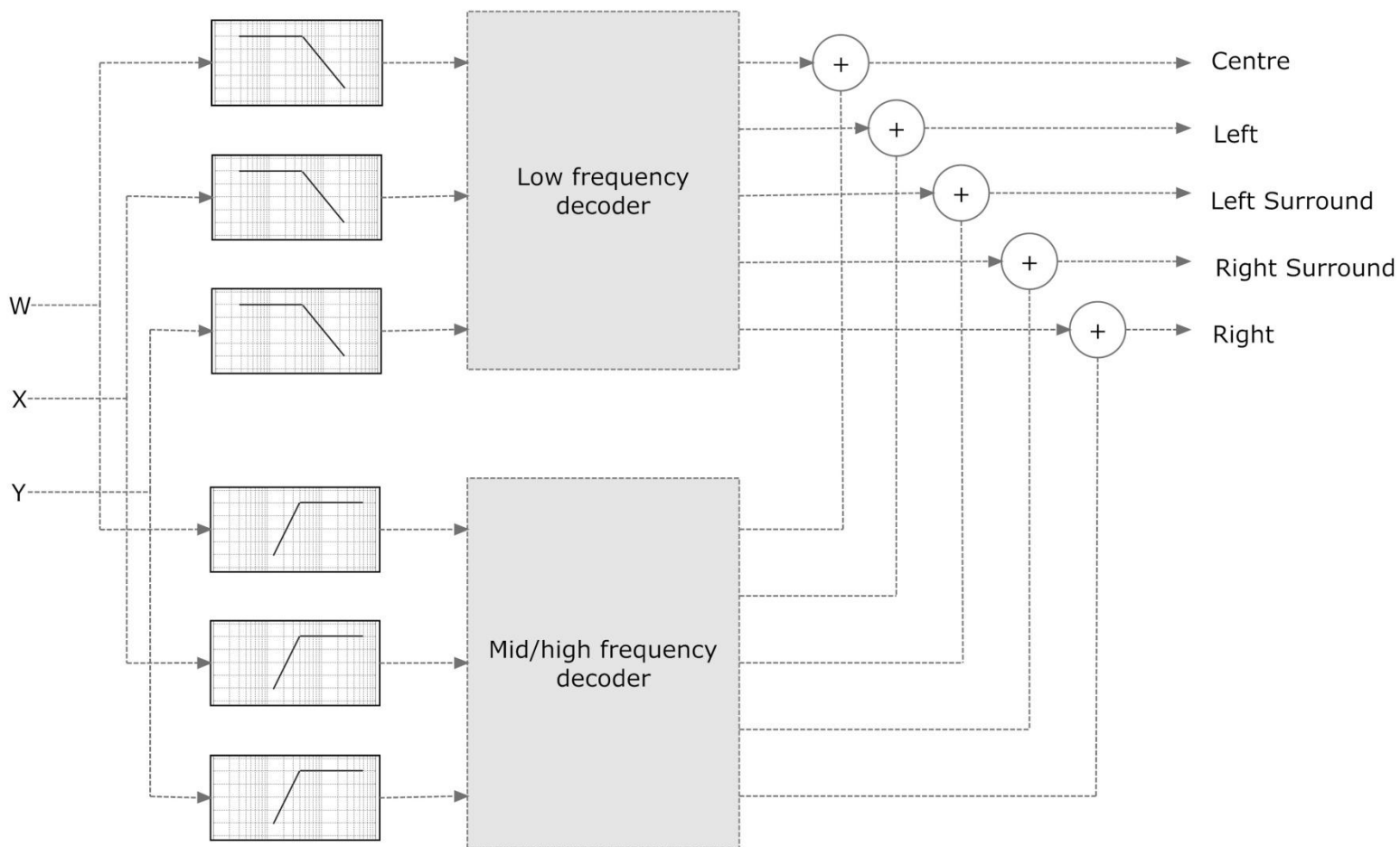


Figure 3-7: Schematic diagram of a first order dual band Ambisonic decoder for irregular loudspeaker arrays

Another important factor to consider when developing Ambisonic decoders is the order of the system. As previously mentioned, the higher the order of the system the more information about a sound field it can describe. For regular decoders (i.e. decoders derived for a regular arrangement of loudspeakers) it is recommended that a minimum number of loudspeakers K are used for a given system order N (see equation 3.16).

$$K = 2N + 1 \quad (3.16)$$

This condition is recommended so the best performance can be achieved in terms of spatial perception and also sound field reconstruction (Ward & T.D. Abhayapala 2001).

For irregular decoders, the above condition does not necessarily yield the best performance. For irregular decoders it is possible to use higher order components to optimise the virtual microphone response to better fit the arrangement of loudspeakers. For example, a decoder derived for the ITU array could use a mix of higher order components in the front of the system (where the loudspeakers are closer together) and lower order components in the rear of the system (where the loudspeakers are further apart). Previous work by others has shown this to be an effective technique (Craven 2003; Wiggins 2007; Poletti 2007). In the literature the terms ‘panning law’ and ‘higher order decoder’ have been used to describe these system. In this thesis we use the latter.

The Ambisonic system is adopted for use in this research for the following reasons:

1. Ambisonics recognises the fact that human hearing uses different mechanisms for localising sound in different frequency regions. It is built around respected theories of auditory localisation (i.e. the velocity and energy models).
2. Research has shown that Ambisonics has a larger sweet spot than other techniques commonly used for ITU 5.1 playback (Bamford & Vanderkooy 1995). Furthermore, the size of the sweet spot can also be increased by using higher order systems (Malham 1999; Daniel & Moreau 2004).

3. Recent work by Wiggins has developed a novel method for optimising irregular Ambisonic decoders (Wiggins 2004). There is scope for developing this work further.

3.4 Tabu search

When there are multiple parameters involved in a search problem searching exhaustively for the best solution is not always feasible. In this work the decoder coefficient search space is large. For example, searching for a first-order frequency independent decoder for the ITU 5-speaker layout (8 decoder coefficients) using a resolution comparable with currently published decoders (4 decimal places) would involve evaluating 10^{32} potential solutions i.e.

$$1 / 0.0001 = 10^4$$

$$10^4 \times 10^8 = 10^{32}$$

Consequently, it is impractical to undertake an exhaustive search of all possible sets of decoder values. In the absence of any *a priori* information being available to reduce the range of valid coefficient values, a local search algorithm must be used to attempt to find good solutions.

This work will use the Tabu search algorithm for finding good decoder coefficients as evaluated by the fitness functions which were developed as part of this research. It must be noted that when a heuristic search algorithm like the Tabu search is used there is no guarantee the globally best solution will be found. However, given enough search runs they should almost certainly provide a good solution.

The Tabu search explores a search space with the aim of finding the best solution possible. The algorithm is “intelligent” in that it enhances its performance by using memory structures. One of these memory structures is known as the Tabu list - a list of previous moves which are designated out-of-bounds, or Tabu (hence the name). The Tabu list is used to guide the search away from previously visited areas in the search space preventing search cycling (search cycling is where the algorithm gets stuck in a local minimum of the search domain) (Glover 1989; Glover 1990).

Moving away from local minima in this manner increases the likelihood of finding a better area in the search space and thus increases the potential of finding a better solution. Figure 3-8 describes the Tabu search algorithm:

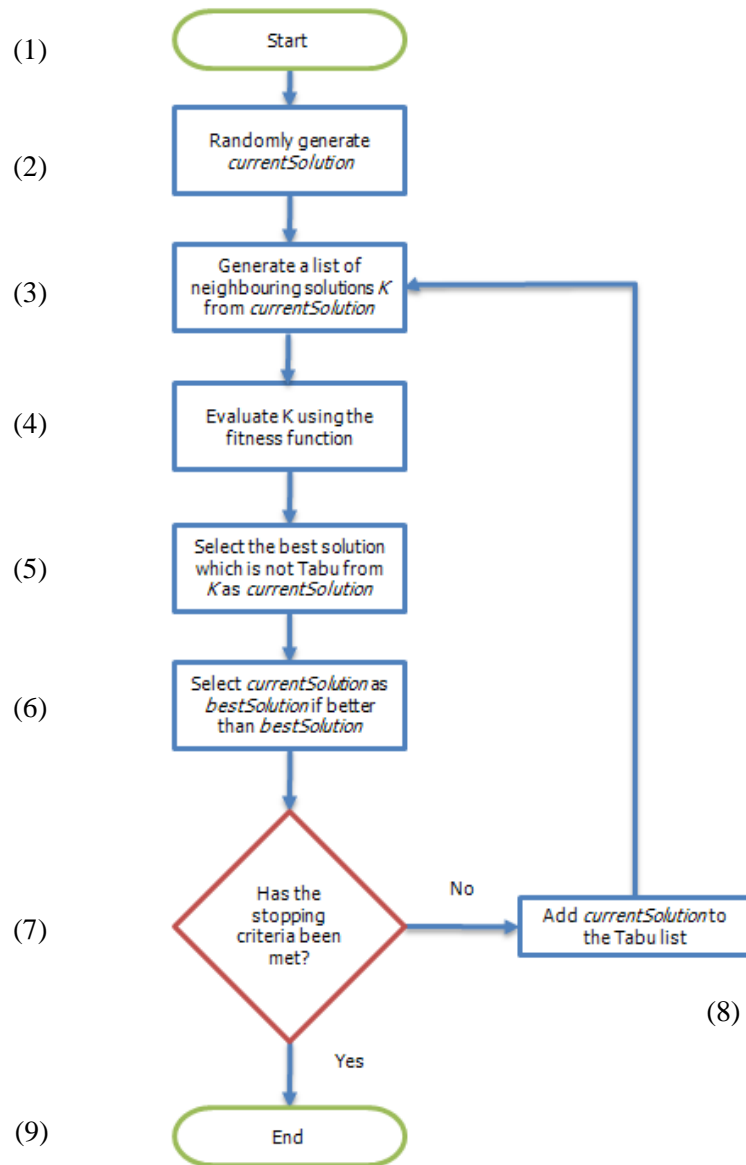


Figure 3-8: The Tabu search algorithm

When generating the neighbouring solutions (step 3 of the Tabu search algorithm) it is possible to use a number of different move types. Usually, the move type is problem-dependent and tailored specifically to match the needs of the problem. For example, in the well-known chessboard optimisation problem known as N-Queens a swap move is used to swap the positions of the queens on the chessboard. In this work the approach will be to step in positive and negative directions along each coordinate axis in the search space (the coordinate axes correspond to the decoder coefficients). This move type allows the search to iterate through all possible local solutions to a set resolution. A fixed step size of 0.0001 will be used as this resolution is comparable with previously published decoders (Gerzon & Barton 1992).

Each of the neighbouring moves generated in step 3 is evaluated by a fitness function with the search algorithm selecting the move with the best fitness score (step 4 of the algorithm). This process is repeated starting from the newly selected current best point in the search spaces until the stopping criteria has been met.

Different stopping criteria can be used. One method is to stop the search after a fixed number of moves (as used by Wiggins). The advantage of this using this approach is the search is guaranteed to stop within a set amount of time. The disadvantage, however, is the search could stop before reaching a minimum in the search space. Another method is to stop the search after a specific goal has been reached in terms of solution fitness. This approach is suitable if the user has a minimum requirement for a solution's quality. However, the search is potentially giving up on finding much better solutions. In this work, the search will be stopped after a fixed number of *bad* moves have been made. This allows the search potentially to reach better solutions when compared to stopping after a fixed number of moves.

Additional stages can be added to the Tabu search algorithm if required. These include a diversification stage where the algorithm explores different areas of the search space if solutions around the current area are deemed poor. An intensification stage can also be incorporated where the algorithm intensifies its search in the area where the best solutions were found (Hertz et al.

n.d.). In this work, however, these criteria will not be used because there is no guarantee that implementing these extra stages will yield better solutions than when just running the basic Tabu search algorithm multiple times.

The Tabu search will be used for producing decoder coefficients according to fitness functions developed in this research. This particular algorithm was chosen as it is a good heuristic search algorithm which performs consistently in multi-objective optimisation. Furthermore, it is tried and tested in this line of work (Wiggins et al. 2003; Wiggins 2004; Wiggins 2007). Although other search algorithms were tested in the initial stages of this research (e.g. Simulated Annealing and a Genetic algorithm) none were able to produce better solutions or produce solutions more quickly than the Tabu search.

3.5 Summary

This chapter has presented background theory on the three techniques chosen for use in this research: the velocity vector and energy vector, the Ambisonic system and the Tabu Search. It has also shown their use in this work.

Chapter 4

A Software Based Design Tool for Producing 5-Speaker Surround Sound Decoders

4.1 Introduction

The aim of this research was to produce a flexible software-based tool for designing improved surround sound decoders. The finished tool provides the user with a high-level interface for executing a search for decoder parameters that best fit the fitness function criteria developed in this research (the user interface is show in figure 4-1). By adjusting the interface controls the user can produce decoders with different performance characteristics.

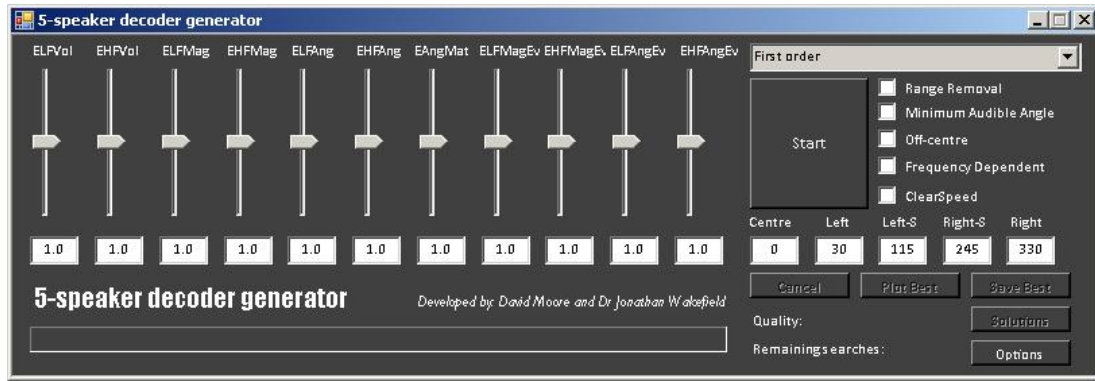


Figure 4-1: Software-based decoder design tool

Each section of this chapter describes a component of the design tool. In **section 4.2** the main multi-objective fitness function algorithm used for guiding the Tabu Search is presented. This algorithm was specifically designed to improve upon previously published work. **Section 4.3** will detail two techniques known as range-removal and importance. Range-removal was introduced to resolve the problem of certain fitness function objectives dominating the search. Importance was introduced to allow the user to logically bias range-removed objectives. **Section 4.4** discusses the option for allowing the user to vary the Ambisonic decoder order. **Section 4.5**

describes a method added to the design tool for reducing the localisation performance variation of decoders around the 360° sound stage. This variation in performance is inherent in all previously published Ambisonic decoders for irregular 5-speaker layouts. **Section 4.6** describes a method for biasing the performance of decoders in directions of the sound stage where humans are more sensitive to sound localisation. **Section 4.7** describes a method that allows the localisation performance of a decoder to be optimised for off-centre listeners. **Section 4.8** explains how the design tool takes advantage of today's High Performance Computing hardware to accelerate the search process. Finally, the penultimate section of this chapter, **section 4.9**, details each of the tool's user interface controls.

4.2 Improved multi-objective fitness function

The multi-objective fitness function used for guiding the Tabu Search encapsulates criteria from the velocity and energy models. The input to the function is a set of decoder parameters generated by the search which are used to determine the amount of Ambisonically encoded audio played out of each loudspeaker (see chapter 3).

The fitness function algorithm is based on an algorithm implemented by Wiggins (Wiggins et al. 2003) and involves checking multiple objectives at equally spaced angles around one side of the left/right symmetrical ITU sound stage. The function builds upon Wiggins' work and aims to match Gerzon's specification for the Ambisonic system more closely.

4.2.1 Volume objectives

When developing a decoder for an irregular array of loudspeakers it is important to ensure the perceived volume is equal all the way around the listener at low and mid/high frequencies. This is because in an irregular loudspeaker array there will be a greater concentration of loudspeakers in certain regions of the 360° sound stage. Consequently, if the same magnitude is used for each virtual microphone response, the overall volume will be louder where there are a greater number of loudspeakers, and quieter where there are fewer loudspeakers.

The volume objectives proposed by Wiggins compare the volume at every angle against the volume at zero degrees. However, this does not necessarily find solutions where the difference between the volume at each angle is similar. In this work the volume at every angle is compared to the volume at all other angles to ensure the error is reduced (see equations 4.1 and 4.2). The reader is reminded that at low frequencies the pressure P is used to represent the perceived volume for the listener, whereas at mid/high frequencies the energy E is used.

$$E_{LFVol} = \frac{1}{n^2} \sum_{i=0}^{180} \sum_{j=0}^{180} |1 - P_i/P_j| \quad (4.1)$$

$$E_{HFVol} = \frac{1}{n^2} \sum_{i=0}^{180} \sum_{j=0}^{180} |1 - E_i/E_j| \quad (4.2)$$

where E_{LFVol} is the absolute error difference of the pressure, E_{HFVol} is the absolute error difference of the energy, P_i and P_j are the pressure at i and j degrees respectively, and E_i and E_j are the energy at i and j degrees respectively. When E_{LFVol} and E_{HFVol} are equal to zero this equates to a constant volume level as a source is panned around the listener.¹

4.2.2 Vector angle objectives

Gerzon (Gerzon 1980) states that the velocity and energy vector angles will coincide if the following three conditions are met:

1. All speakers are the same distance from the centre of the layout
2. Speakers are placed in diametrically opposed pairs
3. The sum of the two signals fed to each diametric loudspeaker pair is the same for all diametric pairs

¹ Please note that these volume objective equations were used when deriving decoders in this thesis. However, more computationally efficient versions of these equations are described in Appendix A.

Only the first of these conditions is met with an irregular ITU 5-speaker decoder so it can be taken that the localisation vectors will not coincide. The following objectives are proposed to ensure this performance error is minimised for each angle θ around the central listening point:

$$E_{LFAng} = \sum_{i=0}^{180} |\theta_i^{Enc} - \theta_i^V| \quad (4.3)$$

$$E_{HFAng} = \sum_{i=0}^{180} |\theta_i^{Enc} - \theta_i^E| \quad (4.4)$$

where E_{LFAng} is the error between velocity vector angle and the desired encoded source angle and the, E_{HFAng} is the error between the energy vector angle and the desired encoded source angle, θ_i^{Enc} is the encoded source angle at i degrees, and θ_i^V and θ_i^E are velocity and energy vector angle at i degrees respectively.

4.2.3 Angle match objective

When applying the velocity and energy localisation vectors to decoder design Gerzon states that it is important for the vector angles to match up to around 4 kHz (Gerzon & Barton 1992). In the fitness function implemented by Wiggins, this important point was not included. Aiming to match the encoded source angle with the velocity vector angle and the encoded source angle with the energy vector angle does not necessarily ensure the angle between the two localisation vectors is minimised. Consider the following two examples (A and B) in figure 4-2. In both examples the error according to the vector angle objectives (E_{LFAng} and E_{HFAng}) is the same. However, example B would be more desirable than example A according to Gerzon's requirements because the velocity vector angle (θ^V) and the energy vector angle (θ^E) are a closer match.

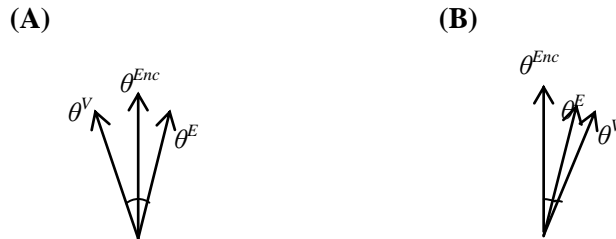


Figure 4-2: Vector angle match problem

To address this issue, a further objective was specifically designed to ensure the error between both vectors is minimised:

$$E_{AngMatch} = \sum_{i=0}^{180} |\theta_i^V - \theta_i^E| \quad (4.5)$$

where $E_{AngMatch}$ is the error between the velocity and energy vector angles, θ_i^V and θ_i^E are velocity and energy vector angle at i degrees respectively.

4.2.4 Vector magnitude objectives

As previously highlighted, a localisation vector length of 1 is optimum. Therefore, the aim of the following vector magnitude objectives is to minimise the error at each angle between the ideal length and the reproduced length:

$$E_{LFMag} = \sum_{i=0}^{180} |r_i^{Enc} - r_i^V| \quad (4.6)$$

$$E_{HFMag} = \sum_{i=0}^{180} |r_i^{Enc} - r_i^E| \quad (4.7)$$

where E_{LFMag} is the error between and ideal velocity vector length ($r_i^{Enc} = 1$) and the reproduced velocity vector length, E_{HFMag} is the error between the ideal energy vector length ($r_i^{Enc} = 1$) and the reproduced energy vector length, r_V and r_E are the magnitudes of the velocity and energy vector at i degrees respectively.

4.2.5 Implementation details

All objectives were designed to be computationally efficient because the fitness function will be called many times by the search algorithm. For example, taking the absolute value of the objective error was preferred to the root mean square method previously suggested by Wiggins to

reduce computational complexity. The following table describes the fitness function algorithm as a whole using pseudo code.

<p>FOR each sound source angle</p> <p> CALCULATE loudspeaker gains</p> <p> CALCULATE pressure</p> <p> CALCULATE energy</p> <p> CALCULATE velocity vector</p> <p> CALCULATE energy vector</p> <p> CALCULATE each fitness function objective and ACCUMULATE their values</p> <p>ENDFOR</p> <p>SUM the fitness function objectives to obtain the total fitness</p>

Table 4.1: Core multi-objective fitness function algorithm described using pseudo code

Some of the calculations in the fitness function require additional information. For example, when computing the loudspeaker gains, knowledge of the encoding gains is required. Likewise, when computing the velocity vector and energy vector, knowledge of the loudspeaker angles is required. In order to maximise efficiency, each of these additional factors was calculated only once in an initialisation stage prior to the start of each search.

4.2.6 Evaluating frequency dependent decoders

As highlighted in chapter 3, Ambisonic decoders for irregular loudspeaker arrays can use separate sets of parameters for low and mid/high frequencies so both the velocity vector and energy vector can be optimised. When evaluating such decoders in this work the two separate sets of parameters were combined and evaluated by the improved fitness function.

4.2.7 Summary

This section has described the improved multi-objective fitness function used for guiding the Tabu Search. The individual objectives that make up the function match the requirements of the Ambisonic system more closely than in previous work. Specifically, objectives E_{LFFol} and E_{HFFol} are improvements to Wiggins' objectives to more closely match his intentions, whereas $E_{AngMatch}$ is a new objective added to more closely match Gerzon's definition of an Ambisonic system. The fitness function was designed to be as computationally efficient as possible.

This algorithm formed the basis on which subsequent components of the design tool were built.

4.3 Range-Removal and Importance

During early testing of the improved fitness function a deficiency was identified with aiming to meet multiple objectives simultaneously (David Moore & J. P. Wakefield 2007). The crux of the problem lies in each of the fitness function objectives having a different numerical range. This effectively biases a search in favour of the objectives with the largest range, causing them to dominate the search and become better optimised at the expense of other objectives.

In order to address this problem, a technique known as “range-removal” was introduced into the optimisation process to systematically and logically remove this bias. A further technique termed “importance” was also introduced for biasing range-removed objectives (David Moore & J. P. Wakefield 2007). This section describes these two important techniques. It should be noted that all previously published work in this application area has not addressed the problem of objective dominance apart from by *ad hoc* objective weighting.

4.3.1 Objective dominance

In order to explain the problem of objective dominance, consider the following two abstract objectives which are to be *minimised* by a search. Objective one represents low frequency localisation quality, which for the sake of argument, ranges from 0 to 10,000. Objective two represents high frequency localisation quality, which ranges from 0 to 2. If the objectives were simply summed to obtain a total fitness value it would be easier for the search to produce solutions with better performance for objective one. For example, if the value of objective two were to decrease from 50%, the consequence would be *insignificant* in the terms of total fitness (i.e. $\approx 0.01\%$). However, if objective one were to decrease by 50% the consequence would be *significant* in terms of total fitness (i.e. 49.99%). Hence, when summing the objectives to obtain a single fitness value, the objective with the largest range (i.e. objective one) would dominate the search.

It is possible to compensate for objective dominance by applying *ad hoc* weightings to individual objectives. However, this is not a satisfactory approach. With the previous example, suppose that the weighting $w_1 = 1$ was applied to objective one and $w_2 = 2$ was applied to objective two. Given the range of the two objectives, objective one would still dominate the search and the use of w_2 would be irrelevant. This highlights a fundamental deficiency in this method of correcting dominance - it can be difficult to discern between setting weights to compensate for differences in objective ranges, and setting weights to indicate the relative importance of an objective.

To demonstrate objective dominance in the context of the current research, table 4.2 displays the mean, minimum value, maximum value and the range of the individual improved fitness function objectives which were recorded over a series of search runs. The values presented are for a typical first-order frequency-independent Ambisonic decoder with ITU surround speakers angled at $\pm 115^\circ$.

Objective:	Mean	Min	Max	Range
E_{LFAng}	0.0012	0.0000	557.0200	557.0200
E_{HFAng}	26.1330	0.0538	529.7900	529.7362
$E_{AngMatch}$	26.1330	0.0586	522.4900	522.4314
E_{LFMag}	0.4956	0.4956	5000000.0000	4999999.5044
E_{HFMag}	62.5960	0.5691	152.7200	152.7200
E_{LFFVol}	0.0000	0.0000	77266.0000	77266.0000
E_{HFFVol}	0.3803	0.0000	326.4800	326.4800
Total	115.7391	-	-	-

Table 4.2: Approximate ranges of the improved fitness function objectives. The objectives with the largest ranges (highlighted) are likely to dominate the search.

It can be seen that the mean values vary substantially. The best mean values are achieved for the low frequency objectives (E_{LFAng} , E_{LFMag} , E_{LFFVol}). All three have significantly lower values when compared with the other objectives and account for less than 1% of the total fitness error (sum of the objectives). The good results for the low frequency objectives and the poor values (in comparison) for the others imply that the low frequency objectives dominate the search for decoder coefficients. This hypothesis is further strengthened by observing the fact that the range

for the low frequency volume objective (E_{LFFVol}) and the low frequency magnitude objective (E_{LFFMag}) is significantly larger than the other objectives.

Interestingly, the range for the low frequency angle objective (E_{LFFAng}) is comparable with the other angle objectives. However, on average, this objective was much closer to its ideal value. The reason for this is likely to be due to objective inter-dependency (i.e. better performance for E_{LFFVol} and E_{LFFMag} led to better performance for E_{LFFAng}).

Despite performing badly, the low minimum values for the high frequency objectives (E_{HFFAng} , E_{HFFMag} , E_{HFFVol}) show that significantly better values can be achieved. This highlights the importance of including a systematic method of objective range-removal in the design tool to help regulate the contribution of each fitness function objective.

4.3.2 Range-removal

Objective range removal is not, in itself, a new concept. Bentley and Wakefield have addressed this generic issue in search problems (Bentley & J. P. Wakefield 1998). The range-removal method used in this application domain comes from their work and is known as the “sum of global ratios”. In this method each of the objective values is converted into a ratio by using the globally worst and best objective values encountered in all previous searches. This ensures that no single objective dominates the search because all values are constrained within the range of [0, 1]. Each objective ratio can be formulated thus:

$$F_i^{Ratio} = \frac{F_i(x) - F_i^{min}}{F_i^{max} - F_i^{min}} \quad (4.8)$$

where F_i^{Ratio} is the i th range-removed objective and F_i is the value of the i th objective given the solution x . F_i^{min} is the minimum value of the i th objective (i.e. the best objective value encountered in all previous searches). Whereas F_i^{max} is the maximum value of the i th objective (i.e. the worst objective value encountered in all previous searches).

Although several other range-removal methods are defined in the literature, this technique was incorporated into the decoder design tool as it has been shown to be robust in a number of different multi-objective optimisation studies (Bentley & J. P. Wakefield 1998; Marler 2005; Marler & Arora 2004).

4.3.3 Importance

Once range-removal has been implemented, a search can be systematically and logically biased towards specific criteria by placing more or less emphasis on selected objectives. This technique is referred to as importance and simply involves applying weightings to the range-removed objectives:

$$F_i^{Weighted} = w_i F_i^{Ratio} \quad (4.9)$$

where $F_i^{Weighted}$ is the i th importance weighted range-removed objective, F_i^{Ratio} is the i th range-removed objective, and w_i is the importance weighting for the i th range-removed objective.

As highlighted earlier, importance weighting can be applied to objectives without range-removal. However, selecting appropriate importance weightings is considerably more difficult when the effective range of the individual objectives is unknown.

4.3.4 Implementation details

Table 4.3 describes how range-removal and importance were incorporated into the improved fitness function algorithm of the design tool.

<p>CALCULATE Improved Fitness Function (see algorithm defined in table 4.1)</p> <p>FOR each fitness function objective</p> <p> IF objective IS GREATER THAN objectiveMax objectiveMax EQUALS objective END IF</p> <p> IF objective IS LESS THAN objectiveMin ObjectiveMin EQUALS objective END IF</p> <p> APPLY range-removal to each objective using current objectiveMin and objectiveMax values</p> <p>ENDFOR</p> <p>MULTIPLY range-removed fitness function objectives with importance weightings</p> <p>SUM the weighted range-removed fitness function objectives to obtain the total fitness</p>
--

Table 4.3: Improved fitness function algorithm with range-removal and importance

In order to derive the objective ratios the minimum and maximum values of each objective were dynamically updated and saved during each search. By continuously updating the minimum and maximum objective values in the search, the approximation of each objective's range steadily improves.

4.3.5 Summary

Range-removal was incorporated into the decoder design tool in order to overcome the problem of objective dominance observed during early testing of the improved fitness function. This technique allows each of the objectives to have an equal impact in the search. Another concept known as importance was also introduced to allow the logical biasing of range-removed objectives.

4.4 Optimisation of higher order decoders

In order to increase the capability of the design tool, a further feature was added to allow the user to derive Ambisonic decoders of different orders. Users of the tool can select from first order decoders up to fourth order decoders. This was an important addition as research has shown higher order decoders can yield better performance for a given loudspeaker array (Craven 2003; Wiggins 2007; Poletti 2007).

When deriving higher order decoders for horizontal 5-speaker layouts, five additional decoder coefficients are required per system order for a frequency independent decoder, and ten additional decoder coefficients for a frequency dependent decoder (see table 4.4). This is significant when the design of a decoder is formulated as a search problem because the size of the search space substantially increases with system order.

	1st order	2nd order	3rd order	4th order
Frequency independent	8	13	18	23
Frequency dependent	16	26	36	46

Table 4.4: The number of decoder coefficients required for decoding over left/right symmetrical 5-speaker layouts (frequency dependent and independent)

The advantage of being able to derive higher order decoders from a designer's point of view is that localisation performance (according to the velocity vector and energy vector) can be improved considerably (see chapter 3). When implementing a higher order decoder, however, more audio channels are required for the encoded audio tracks which, depending on the system order, could be an issue in terms of storage on present day media (i.e. DVD).

This feature was incorporated with all the developed components of the design tool in order to offer maximum flexibility to the user. In the next chapter decoders of different orders will be analysed.

4.5 Even localisation performance optimisation

One of the positive aspects of Ambisonics is that for regular loudspeaker layouts it treats each direction on the 360° sound stage with equal precedence. This results in the isotropic performance characteristics that listeners would experience in a real sound field. However, this is not necessarily the case for decoders designed for irregular loudspeaker layouts. When analysing decoders published in the literature and decoders produced using the improved fitness function it was clear that performance can vary significantly around the 360° sound stage (David Moore & J. P. Wakefield 2007).

This section describes a method incorporated into the decoder design tool for producing Ambisonic decoders for irregular loudspeaker layouts with more even performance by angle. Even localisation performance is important for any application where the decoder designer wishes to give the listener an isotropic listening experience (rather than the frontal-biased experience normally provided for sound to moving picture). Such decoders would have applications in the playback of surround sound mixes of popular music from DVD-A and SACD and reproduction of electroacoustic soundscapes.

4.5.1 An analysis of a typical first order Ambisonic decoder for the ITU layout

In order to illustrate how performance varies around the 360° sound stage, a typical first order Ambisonic decoder designed for the ITU layout will now be analysed. The decoder was derived using the improved fitness function with range removal incorporated. All fitness function objectives were given equal importance in the search. Figure 4-3 plots each fitness function objective across the 360° sound stage. The volume objectives have been omitted from this figure as their error was negligible. The total fitness (sum of the objectives) by angle is also included.

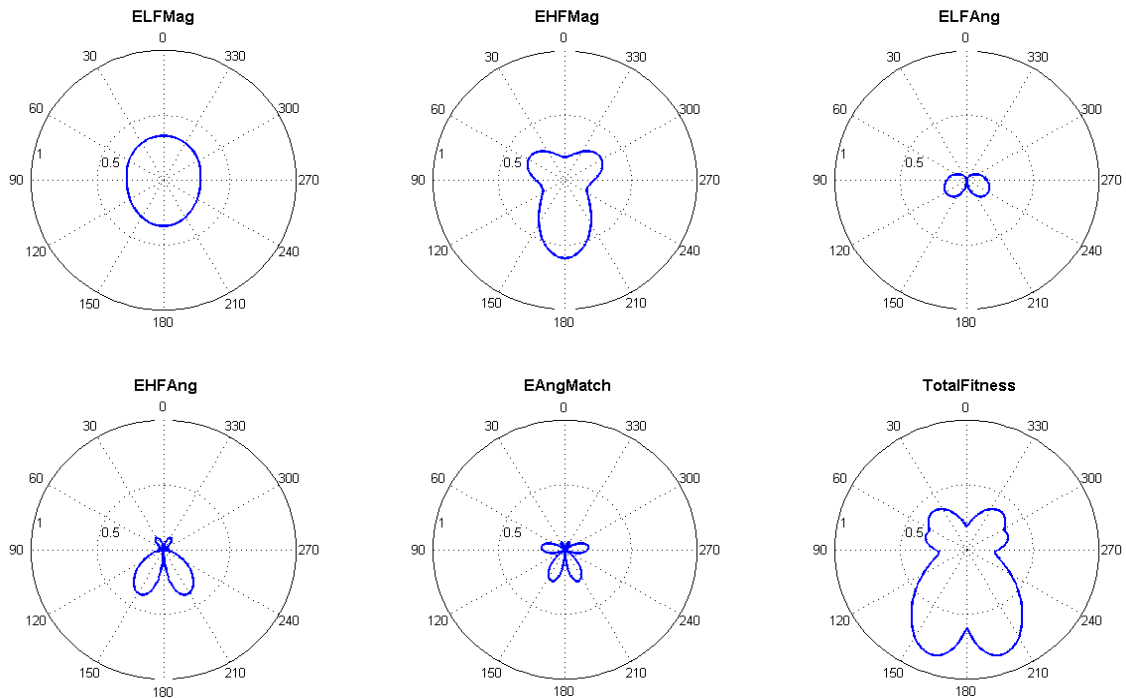


Figure 4-3: The performance of a typical first order frequency independent 5-speaker decoder.

It is clear from figure 4-3 that the response of each objective varies by angle around the 360° sound stage. Generally, objectives are closer to their ideal response at the front and sides of the system rather than at the rear of the system. This is typical of irregular decoders produced using a search because the greatest improvement (in terms of total fitness) can be achieved when maximising performance in the direction of the sound stage with the greater number of loudspeakers. Table 4.5 further highlights the performance variation of this decoder by presenting the standard deviation of the fitness function objective values around the 360° sound stage.

Objective	Standard deviation
E_{LFAng}	0.0615
E_{HFAng}	0.1262
$E_{AngMatch}$	0.0765
E_{LFMag}	0.0241
E_{HFMag}	0.1173
E_{LFVol}	0.0001
E_{HFVol}	0.0001

Table 4.5: Standard deviation of the fitness function objectives values

Clearly, all the objectives for this particular decoder (apart from the volume objectives which were originally designed to ensure even error) have a certain amount of variability. The objectives with the greatest overall variation are the energy vector magnitude and energy vector angle objectives (E_{HFMag} and E_{HFAng}). This large fluctuation in energy vector performance is likely to have a significant impact on the even listening experience for this type of decoder.

In summary of this analysis, the best performance for an ITU 5-speaker decoder is generally in front of the listener, and the worst performance behind the listener (David Moore & J. P. Wakefield 2007; David Moore & J. P. Wakefield 2008). The difference in performance between these two areas is significant in terms of velocity vector and energy vector responses as shown in the above analysis. Moreover, it has recently been shown to be significant when subjectively assessing reproduced audio on these systems (Lee & Hellar 2007).

4.5.2 Even performance design criteria

In order to produce decoders with more even velocity vector and energy vector responses, four additional objectives were incorporated into the improved fitness function. Each uses the standard deviation to measure the performance variation of the vector magnitude (E_{LFMag} and E_{HFMag}) and vector angle objectives (E_{LFAng} and E_{HFAng}) around the loudspeaker layout (see equations 4.10 to 4.13). If the optimum value is met for each of these objectives there will be no deviation from the mean and hence no variation for the corresponding objective.

$$E_{LFAngEven} = \sqrt{\frac{1}{180} \sum_{i=0}^{180} (E_{LFAng} - \overline{E_{LFAng}})^2} \quad (4.10)$$

$$E_{HFAngEven} = \sqrt{\frac{1}{180} \sum_{i=0}^{180} (E_{HFAng} - \overline{E_{HFAng}})^2} \quad (4.11)$$

$$E_{LFMagEven} = \sqrt{\frac{1}{180} \sum_{i=0}^{180} (E_{LFMag} - \overline{E_{LFMag}})^2} \quad (4.12)$$

$$E_{HFMagEven} = \sqrt{\frac{1}{180} \sum_{i=0}^{180} (E_{HFMag} - \overline{E_{HFMag}})^2} \quad (4.13)$$

where $E_{LFAngEven}$, $E_{HFAngEven}$, $E_{LFMagEven}$, $E_{HFMagEven}$ are the standard deviation of the corresponding objectives² defined in section 4.2.

4.5.3 Summary

This section described a method incorporated in the design tool for reducing the large variation in localisation performance by angle around the listening point typically seen in Ambisonic decoders for irregular loudspeaker layouts. The new method uses four new objectives based on the standard deviation. The objectives were specifically designed to reduce the performance variation of the velocity vector and energy vector magnitudes and angles around the 360° sound stage.

4.6 Exploiting human spatial resolution

When reviewing the literature it became clear that the capability of the human auditory system is not often considered when designing surround sound systems (see chapter 2). It appears that most systems, if not all, assume human hearing capability is equal in every direction. However, psychoacoustic research has shown that this is clearly not the case (Blauert 2001; Brian Moore 2003). Humans are more sensitive to sound source localisation in the front and rear than at the sides. After a more detailed look at relevant literature, this section will describe a novel method introduced into the design tool that exploits the resolution of human hearing.

4.6.1 Auditory localisation resolution

The resolution of human auditory localisation can be determined by detecting the smallest noticeable shift in a sound's location. This shift is often referred to as the Minimum Audible Angle (MAA). Work has shown that optimum conditions for the MAA in the horizontal plane are when a sound source is positioned directly ahead of the listener (Mills 1958; Hartmann 1989; Grantham et al. 2003). Under these conditions it is possible to detect shifts of approximately one degree which is generally regarded as the lower limit of auditory spatial resolution (Blauert 2001). Despite being accurate directly ahead of the listener, spatial resolution deteriorates as the

² Please note that a running standard deviation was used when computing each even error objective. The running standard deviation is much more computationally efficient.

source moves to the sides and the rear. Blauert states that spatial resolution at the sides can be between three and ten times worse than at the front and approximately twice as bad at the rear (Blauert 2001). This same pattern of localisation resolution can be seen in the experiments of Mills (Mills 1958), Stevens and Newman (Stevens & Newman 1936), Makous and Middlebrooks (Makous & Middlebrooks 1989) and Saberi et al (Saberi et al. 1991).

There are many other aspects, apart from the direction of the sound source, which directly influence localisation resolution. The frequency content of the sound is important (Mills 1958). Strybel and Fujimoto have shown that the stimulus onset asynchrony (the onset–onset time difference) and the duration of a sound are important (Strybel & Fujimoto 2000). Head movements are important for enhancing spatial acuity (Makous & Middlebrooks 1989; Thurlow & Runge 1967). Furthermore, Chandler and his colleagues demonstrated that *a priori* knowledge of a sound source’s location can aid the listener (Chandler et al. 2005).

Table 4.6 details the stimuli, stimuli duration and number of subjects used in several MAA experiments (please note the different parameters in each experiment). Each of the experiments was undertaken in similar acoustic spaces (i.e. anechoic or treated listening chambers) with the exception of the experiment by Grantham which utilised headphones. The results from each experiment are displayed in Table 4.7 along with a mean MAA for the front and sides.

Author	Stimuli	Duration (ms)	Number of subjects
Mills	Sine 500-750Hz	1000	3
Makous	Band limited noise (1.8 – 16kHz)	150	6
Hartmann	Sine 500Hz	1000	3
Grantham (a)	Wideband noise	300	6
Grantham (b)	High pass noise	300	6
Grantham (c)	Low pass noise	300	6
Perrott	Click train 400Hz	50	4
Saberi	Noise bursts	250	3
Heffnet	Noise bursts	100	4

Table 4.6: Stimuli, duration and the number of subjects from a number of MAA experiments

Author	0°	10°	20°	30°	40°	45°	50°	60°	70°	75°	80°	90°	100°
Mills	1°			1.7°		2°		3.5°		8°			
Makous	2.3°	3.5°	3.9°	4.8°	6°		6.5°	7.5°	7°		8.5°		9.5°
Hartmann	0.9°												
Grantham (a)	1.6°												
Grantham (b)	1.6°												
Grantham (c)	1.5°												
Perrott	0.97°												
Saberi												≈5°	
Heffnet	1.3°			2.8°				4.4°				9.7°	
Mean MAA	≈2.6°							≈7°					

Table 4.7: Estimated MAA values from the aforementioned experiments (see previous table)

While conducting this review it became clear that there was a reasonable number of studies detailing MAA measurements made in front of the listener, however, there was little work detailing measurements made at the side of the listener and hardly any data for measurements at the rear of the listener. It was also found that the number of subjects used in each of the experiments was relatively small. This prompts the question of whether these results can be considered completely reliable. However, what is clear is that spatial resolution degrades when moving from the front to the side.

4.6.2 MAA optimisation criteria

In all previous work in this application area each of the fitness function objectives has been given equal importance around the 360° sound stage. In this feature of the design tool, however, an angle dependent weighting is applied to the velocity vector objectives (E_{LFMag} and E_{LFAng}) and energy vector objectives (E_{HFMag} and E_{HFAng}) to bias their performance in directions which human sound source localisation is more sensitive.

The basic principle is to divide the sound stage into 3 areas: the front (0° - 59°), the sides (60° - 119°) and the rear (120° - 180°). In each of these areas the objectives are assigned a weighting that reflects the importance of localisation accuracy in that area (i.e. the front is given the highest weighting followed by the rear and then the sides which will be given the lowest weighting). In

this work the weightings for each area are inversely proportional to mean MAA in the same area i.e.

$$w = 1 / |MAA| \quad (4.14)$$

where w is the weighting and $|MAA|$ is the mean MAA in the corresponding area of the sound stage (i.e. front, side or rear). Table 4.8 gives the weightings which were incorporated into the design tool. The front and the side weightings were calculated using the data from table 4.7. A rear weighting was chosen based on the front and side weightings.

	Weighting
Front	1
Side	0.1428
Rear	0.5000

Table 4.8: MAA objective weightings

It should be noted that a greater angular resolution is possible when applying the weighting scheme. The reason for dividing the sound stage in such a coarse manner in this work was because of the lack of MAA data in the literature.

4.6.3 Summary

In this section a novel weighting scheme was introduced that was designed to optimise the localisation performance of decoders in directions where human sound localisation is more sensitive. The scheme used a MAA optimisation paradigm where each of the improved fitness function objectives was weighted more heavily in directions of the 360° sound stage with a lower MAA value. The aim of the new method was to provide the user with the option of producing decoders with localisation performance that more closely matched human spatial resolution.

4.7 Optimisation of decoders for off-centre listeners

There has been much discussion in the literature about improving the localisation performance of Ambisonic systems at the sweet spot (Gerzon & Barton 1992; Wiggins et al. 2001; Wiggins et al.

2003; Neukom 2006). However, few studies exist which look at improving the localisation performance of Ambisonic systems in off-centre listening positions. There is clearly a need for research in this area as many systems will be used for playing sound to a distributed audience (especially when set up in a large listening space such as a cinema or auditorium). This section describes a method incorporated into the design tool that allows a decoder's localisation performance to be optimised for off-centre listeners.

4.7.1 Background

Arguably the most commonly referenced work on off-centre surround sound is by Malham (Malham 1992). Malham describes informally several personal experiences of using Ambisonics for playback over different large-scale surround sound rigs. One of the major problems he identifies with delivering surround sound in this way is that at non-central listening positions the sound image is drawn towards the nearest loudspeaker. The reason for this is because a listener in an off-centre position will be nearer or further away from some loudspeakers resulting in time differences and level differences between sound waves arriving from each loudspeaker. This leads to the loss of temporal synchronisation of the contributing sound waves and also a sound intensity bias in the direction of the nearest loudspeaker. As a result phantom images can be distorted, or in worst case scenarios, lost completely.

The main perceptual factor behind the breakdown of phantom images in off-centre listening positions is the precedence effect. This effect says the listener will perceive sound as coming from the direction of the earliest arriving wavefront. However, in reality it is not this straightforward. Predicting the impact of the precedence effect in surround sound listening is difficult as it can be influenced by many factors. For instance, Aarts in his paper on time/intensity trading has demonstrated that sound level differences can override temporal differences and ultimately the precedence effect (Aarts 1993). In addition, a number of studies have shown that the characteristics of the audio signal can directly change the perceptual thresholds in which the precedence effect operates (e.g. the auditory system appears to have less susceptibility to short transient sounds than continuous signals). These factors all play a part in how much room a listener has to manoeuvre away from the sweet spot before the image becomes completely biased towards the nearest loudspeaker.

Malham identified another problem specific to off-centre Ambisonic playback. He observed that first order Ambisonic decoders (designed according to one of Gerzon's theorems) had poor localisation performance in off-centre positions. He noted the reason for this was because first order decoders play sound out of all loudspeakers simultaneously. As a result of this, listeners in off-centre listening positions perceived what Malham terms a "bounce back" effect where sound would effectively be heard in two different locations. In order to remove this effect, Malham later devised the Cardioid decoder where the secondary lobe of the virtual microphone polar response is removed (see chapter 3). However, although this decoder removes the problems of bounce back, it leads to a significant decrease in overall localisation performance at the sweet spot. For example, studies have reported Cardioid decoders as having poor localisation performance with sound images sounding too diffuse (Benjamin et al. 2006; Guastavino et al. 2007).

Recent work by Poletti has introduced a different method of improving the performance of surround sound systems away from the centre point. Poletti's work involves using a least-squares pressure matching method for approximating an optimal fourth order decoder for the ITU 5-speaker layout. Basically, the least-squares approach involves matching the sound pressure at several points in the listening area between an ideal soundfield and the decoded soundfield. One of the advantages of this method is soundfields can be analysed over an area rather than a single point. However, although the pressure matching approach is able to produce theoretically robust solutions, it does not take into account what the listener may perceive. In this work a method was incorporated into the design tool for checking what a listener may perceive at different points in the listening area.

At the time of writing this thesis, the work by Poletti is the only work that details the optimisation of surround sound systems for the ITU 5-speaker layout away from the centre point. Thus, there is clearly a need for further work in this area in order to develop and advance this line of research.

4.7.2 Off-centre evaluation criteria

In order for sound localisation performance to be measured in off-centre listening positions, the velocity vector and energy vector were re-formulated. This re-formulation takes into account the fact that the loudspeakers are at different distances to an off-centre listener and also at different angles (figure 4-4 illustrates this).

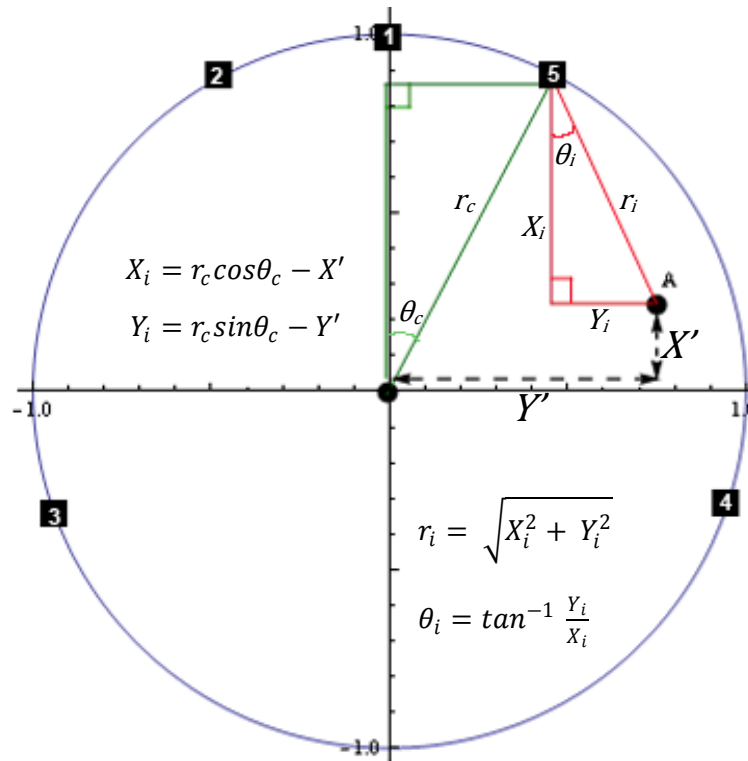


Figure 4-4: The distance and angle of each loudspeaker changes according to the listening position.

It is clear from figure 4-4 that sound arriving at the off-centre position (labelled A) from loudspeakers 1, 5 and 4 will be louder than sound emitted at the same level from loudspeakers 2 and 3. This change in sound level with distance can be modelled using the inverse square law. The inverse square law says that sound intensity decreases as the distance to the source increases i.e.

$$I = \frac{W}{4\pi r^2} \quad (4.15)$$

where I is sound intensity, W is the power of the acoustic source in watts and r is the distance to the source in metres. This is due to the fact that sound energy spreads out as it propagates through the air (Howard & Angus 2001). From equation 4.15 it is clear that every time the distance from a sound source is doubled, sound level intensity reduces by a factor of four obeying the inverse square law:

$$\frac{1}{r^2} \quad (4.16)$$

Because sound level pressure is proportional to the square root of sound intensity, the following equation can be used to model the sound pressure level differences a listener would encounter for each loudspeaker when situated in an off-centre position:

$$g_i = \frac{1}{r_i} \quad (4.17)$$

where g_i is the difference in sound pressure level for the i th loudspeaker and r_i is the distance to the i th loudspeaker. Please note that this equation assumes free field listening conditions (an environment with no reflections). In reality, however, there will be sound interaction with objects and walls in the listening environment. There will also be air temperature fluctuations making sound level changes with distance very complex. Nevertheless, equation 4.17 provides a good first approximation of the change in sound level over distance.

When calculating the pressure, velocity vector, energy and energy vector in an off-centre position this gain factor is directly applied to all loudspeaker gains i.e.

$$S_i = g_i S_i^{Original} \quad (4.18)$$

where $s_i^{original}$ is the loudspeaker signal for the i th loudspeaker. In addition, it is important to include the new angles of the loudspeakers (i.e. θ_i) from the off-centre position in the equations.

Previously, when estimating sound localisation from the centre point in the improved fitness function, the optimum length of the velocity vector and energy vector was unit magnitude. However, the optimal length of both vectors will change according to the distance from the origin, and also the angle of the sound source. The optimum length of each vector when measuring in an off-centre position is equivalent to the distance from the listening position to a sound source on the boundary of the listening area. The optimal vector angles will also be different at each listening position.

4.7.3 Implementation details

Table 4.9 describes how the off-centre optimisation criteria were incorporated into the fitness function. Please note that this algorithm is only concerned with adjusting each of the loudspeaker gains, consequently time delay compensation is considered outside the scope of this work.

```

FOR each listening position

    CALCULATE the angles of the loudspeakers from the current position
    CALCULATE the distance to the loudspeakers
    CALCULATE the loudspeaker gain scaling factors

    FOR each sound source angle

        CALCULATE the ideal vector angle from the current position
        CALCULATE the ideal vector magnitude from the current position
        CALCULATE loudspeaker gains (with scaling factors applied)
        CALCULATE local pressure
        CALCULATE local energy
        CALCULATE local velocity vector
        CALCULATE local energy vector
        CALCULATE each fitness function objective and ACCUMULATE their values

    ENDFOR

    UPDATE each objectiveMax and objectiveMin (see algorithm defined in table 4.3)
    APPLY range-removal and importance
    SUM the fitness function objectives and ACCUMULATE to obtain the total fitness

ENDFOR

```

Table 4.9: Off-centre fitness function algorithm

In this implementation range-removal is position dependent. That is, different minimum and maximum values are stored at each evaluated position to take account of the different possible objective ranges at each position.

Please note that the runtime performance of this algorithm is highly dependent on the number of listening positions checked in the fitness function. Wherever possible, values were pre-calculated before the two main ‘for’ loops to improve runtime performance (as with the original improved fitness function algorithm).

4.7.4 Summary

This section has detailed another component of the design tool – the ability to optimise the localisation performance of decoders in distributed listening positions. This component is important because surround sound is often played to an audience with multiple listeners distributed in the listening area. The method involved re-formulating the velocity vector and energy vector to take into account the different loudspeaker angles and distances the listener would encounter when in an off-centre position. The inverse square law was used to model the sound pressure level changes for the loudspeakers over distance.

4.8 Search Acceleration using High Performance Computing Hardware

The final addition to the design tool was the ability to run searches on High Performance Computing (HPC) hardware. HPC technology is becoming more accessible to general users because of the decrease in price of hardware components, and the increase in network technology performance (El-Rewini & Abd-El-Barr 2005). This is opening up an array of possibilities in different fields of research. For instance, applications previously disregarded as too computationally expensive to compute are being reconsidered.

The term HPC used to refer directly to the work of supercomputers. However, nowadays it encompasses a wide range of computing resources such as: computer graphics processors units (GPUs) with multi-processor core architectures, Hardware Applications Accelerators (e.g. ClearSpeed multi-processor boards), clusters of networks computers and GRIDs (multiple

computing resources connected through the internet). GPUs and Application Accelerators are compact solutions to HPC which can be used in conjunction with desktop computers whereas clusters and GRIDs are distributed computing solutions which potentially require more management. In this work we will use a ClearSpeed Application Accelerator.

The ClearSpeed HPC hardware was used to accelerate the search process of the design tool allowing a greater number of searches to be run within a period of time potentially leading to better solutions being found (David Moore & J. P. Wakefield 2009). Furthermore, it was used so the tool would become more responsive (because of faster search times) leading to a higher level of interactivity with the user.

4.8.1 Implementation details

Two ClearSpeed x620 boards were used for accelerating the search (see figure 4-5). The x620 boards have dual CSX600 chips and 1 GB SRAM. Each chip has an array of 96 processor elements (PE) that each operate at 250 MHz and have 6KB of local memory. The ClearSpeed boards have Single Instruction, Multiple Data (SIMD) architectures where multiple processors simultaneously execute the same instruction but on different data.



Figure 4-5: Clearspeed x620 board

The boards were programmed in a SIMD style using the Cⁿ language (an extension of the C programming language). Cⁿ has special data types to differentiate between nonparallel data

instances (mono) and parallel data instances (poly). ClearSpeed provide optimised standard math functions which process poly-scalars (i.e. one piece of data per PE) or poly-vectors (i.e. 4 pieces of data per PE). Poly-vectors more efficiently exploit the parallel architecture of the boards by allowing 384 calculations to be made simultaneously on each chip (i.e. 96 PEs x 4). An example program illustrating the different data types is provided in figure 4-6.

```
#include <lib_ext.h>
#include <vmathp.h>

// __NUM_PES__ is the number of processor element (96)
#define SAMPLES (__NUM_PES__ * 4)
#define PI 3.14159265358979

int main(void)
{
    // __FVECTOR is a poly-vector
    __FVECTOR sine, angle = {0,0,0,0};

    // get_penum() returns the ID of each PE (0 - 95)
    poly int pnum = get_penum();

    // Set up the angles for each element of the vector
    angle[0] = (__NUM_PES__ * 0 + pnum) * PI / SAMPLES;
    angle[1] = (__NUM_PES__ * 1 + pnum) * PI / SAMPLES;
    angle[2] = (__NUM_PES__ * 2 + pnum) * PI / SAMPLES;
    angle[3] = (__NUM_PES__ * 3 + pnum) * PI / SAMPLES;

    // calculate sine of angle
    sine = cs_sinp(angle);

    return 0;
}
```

Figure 4-6: An example Cⁿ program for the ClearSpeed HPC hardware

The Tabu Search and improved fitness function were coded using the poly-vector data type. By using this data type 4 searches could be run effectively in parallel on each PE leading to a total of 1536 (4 x 384) simultaneously executing searches (i.e. 2 boards each with 2 chips). The design tool connects remotely to a server housing the boards at Bath University in the United Kingdom.

When coding the fitness function the algorithm remained the same. However, when coding the search a few changes were necessary to take advantage of the ClearSpeed boards' architecture.

Specifically, the Tabu list was not implemented because of the limited amount of available memory on each processing element. Also, the search was stopped after a fixed number of moves rather than a fixed number of bad moves. It should be noted that stopping the search after a fixed number of moves is not normally ideal. It is generally considered more appropriate to stop a search after a fixed number of *bad* moves to allow the search to reach a local minimum (this implementation would be better suited to a MIMD architecture). However, on a SIMD architecture this will ensure that all PEs are fully employed because they will all start and end each search at the same time.

The advantage of coding the algorithm in this way is the fact that it is scalable. If there are more ClearSpeed boards available to the user then more searches can be run. For example, if 3 boards were available then 2304 searches could run simultaneously. If 4 boards were available then 3072 searches could be run simultaneously.

4.8.2 Summary

In all previously published work in this application area, searches for decoder coefficients have been run sequentially (Wiggins et al. 2003; Craven 2003; Wiggins 2007). In this work, however, High Performance Computing hardware has been incorporated into the decoder design tool to allow multiple searches to be run concurrently. Incorporating this feature increases the probability of finding a good decoder because significantly more potential solutions can be evaluated within a set time. This will be demonstrated in the next chapter.

4.9 Decoder design tool user interface

The design tool consists of a main user interface and two sub-panels (see figure 4-7). The main user interface is the top level of the application where all of the tool's main functionality can be controlled. The performance panel provides detailed information about decoders produced by the search algorithm and the options panel enables the user to configure properties of the Tabu Search algorithm. The following subsections describe each of the elements in turn.

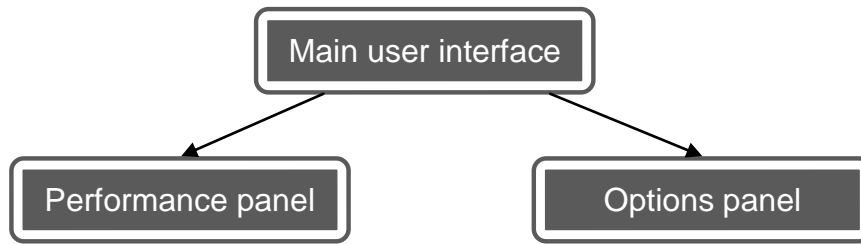


Figure 4-7: Decoder design tool structure

4.9.1 Main user interface

The main user interface (shown in figure 4-8) has a number of different parameters that can be set before starting a search. The user can set importance weightings for each of the improved fitness function objectives by either adjusting the corresponding slider controls or entering values in the edit boxes. In addition, the user can enter the order of the required decoder by selecting from the ‘decoder order’ drop down box (see option 2 in figure 4-8). This drop down box gives the option of deriving decoders from first order to fourth order. Another important feature of the design tool is the ability to switch the developed components on and off in different combinations (see option 3). These components are controlled by check boxes and manage the ability to: apply range-removal, set a minimum audible angle weighting scheme, optimise for off-centre listeners, produce a frequency dependent or independent decoder, run multiple searches in parallel on the ClearSpeed HPC hardware. Element 5 allows the user to load or save solutions produced by the search. When loading a solution the user has the option to use it as the starting point of the search (rather than a random start point). Element 6 allows the user to view a list of all solutions produced by the search from the most recent search run. Finally, the user can input the angles of the loudspeakers using the edit boxes highlighted as Element 4.

4.9.2 Performance panel

When opening the performance panel, the localisation performance of the best decoder produced by the search is detailed (see figure 4-9). There are four plots showing the following information:

1. Plot 1 (highlighted as element 10) shows the velocity vector response around the 360° sound stage. Velocity vector magnitudes are shown in red at each angle and velocity vector angles are displayed as red lines every 30 degrees (starting from 0 degrees at the front of the system). Ideal vector magnitudes and angles are shown in grey.
2. Plot 2 (element 13) shows the energy vector response around the 360° sound stage. Energy vector magnitudes and angles are displayed in green with ideal magnitudes and angles in grey.
3. Plot 3 (element 11) shows the low frequency virtual microphones and pressure.
4. Plot 4 (element 14) shows the mid/high frequency virtual microphones and energy.

Performance plots can be saved as high quality JPEG or PNG image files.

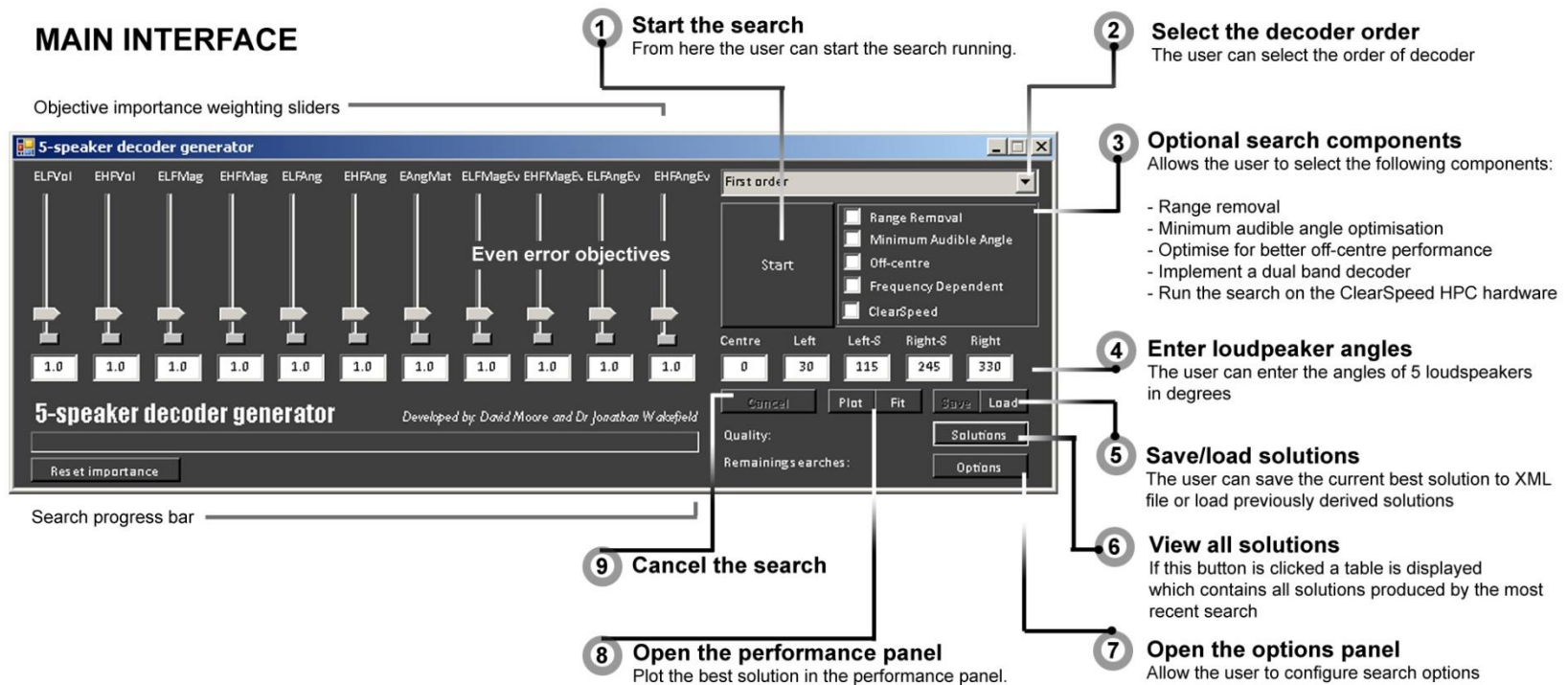


Figure 4-8: Main interface of the decoder design tool

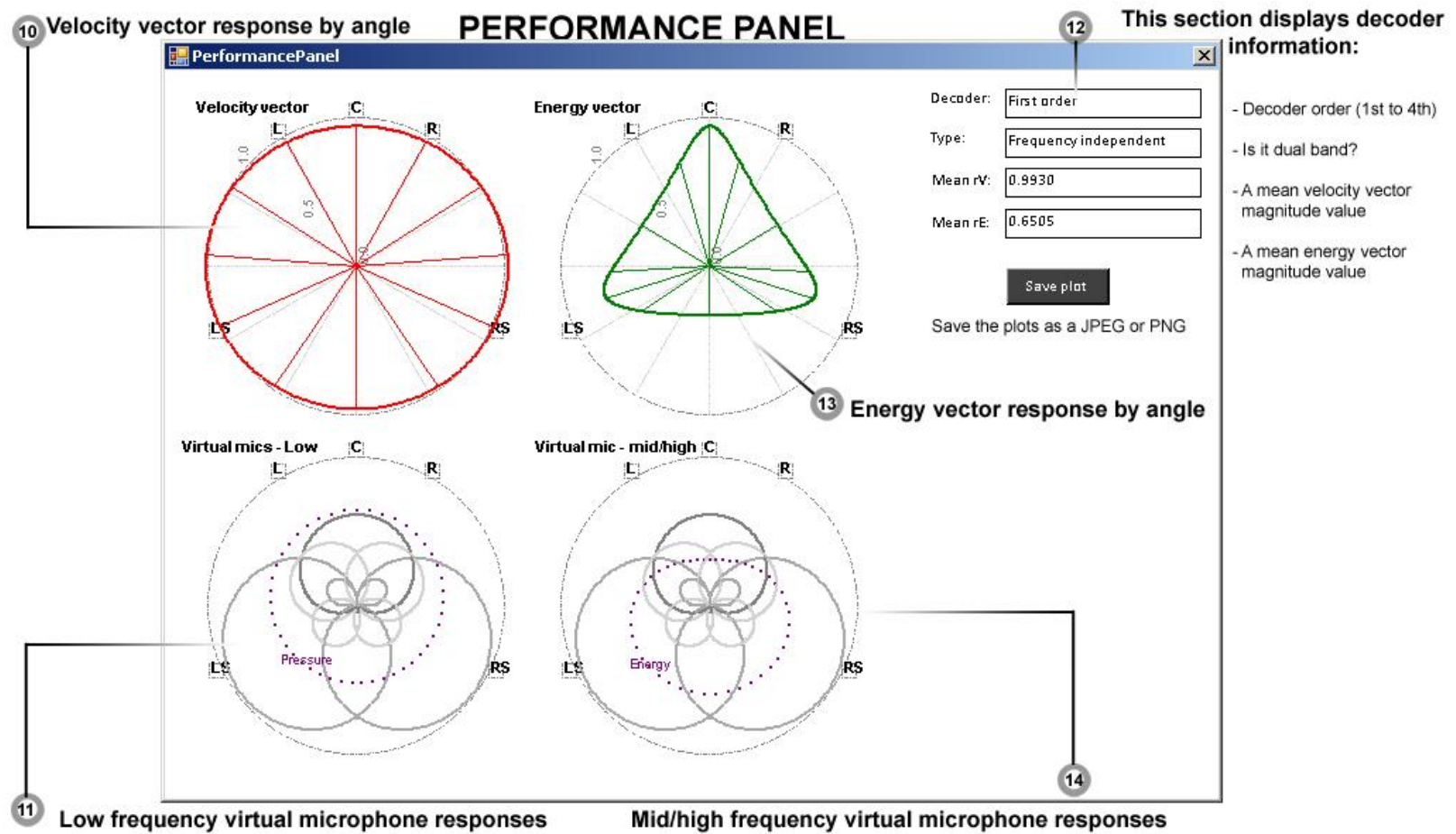


Figure 4-9: Performance panel where the performance of the current best solution produced by the search can be viewed

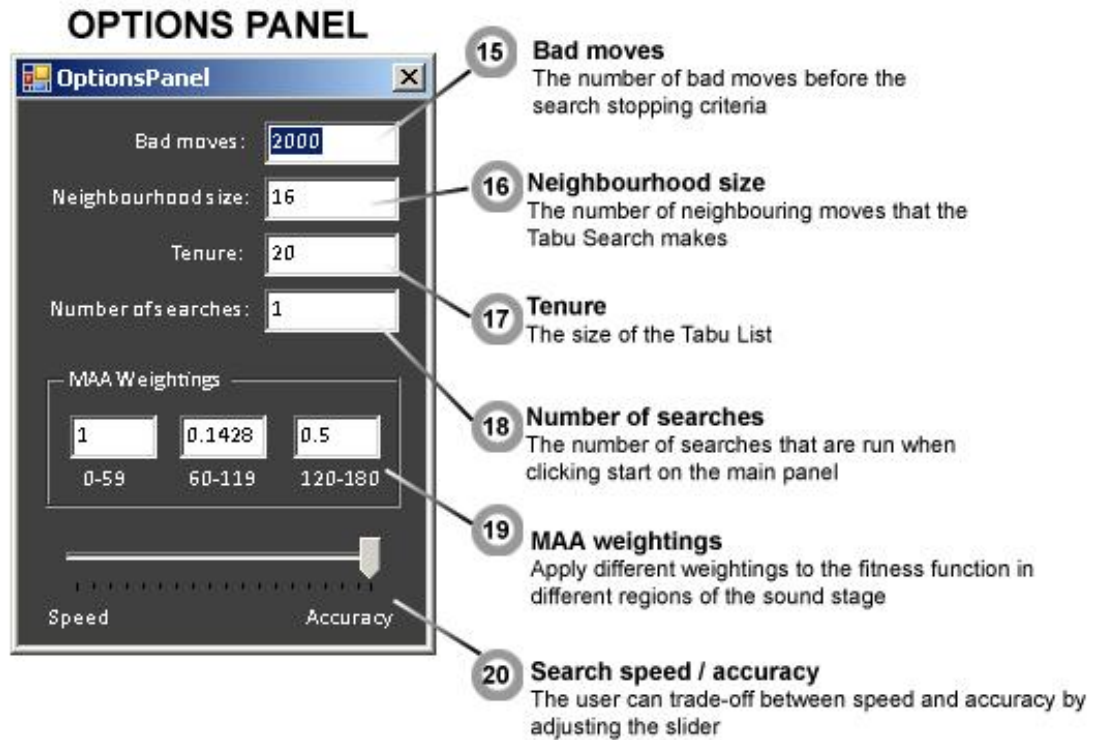


Figure 4-10: Search options panel

4.9.3 Options panel

The options panel allows the user to set the main properties of the Tabu Search (see figure 4-10). By using option 15 of the design tool the user can enter the number of bad moves before the Tabu Search stops running. A higher value for this parameter might lead to a better solution being found as the search could potentially reach a better local minimum, however, a higher number of bad moves is likely to have a direct impact on time-to-solution. Option 16 allows the Tabu Search neighbourhood size to be set. The neighbourhood size is the number of local solutions the Tabu Search generates when searching around the current best solution. The default neighbourhood size is twice the number of coefficients so a positive and negative step can be made for each coefficient. For example, a first order frequency-independent decoder requires 8 coefficients so the default neighbourhood size will be 16. Option 17 allows the user to set the

Tabu Tenure (i.e. the size of the Tabu List). A larger tenure will result in slower search times as the search has to traverse the list for ‘Tabu’ solutions at each iteration of the algorithm. However, a larger tenure will reduce the chance of the search returning to the same local minimum. On the other hand, a smaller tenure will result in the algorithm running faster but may prevent the search from visiting a wider area of the search space. Option 18 allows the user to set the total number of sequentially run searches both on the host machine or the ClearSpeed HPC hardware. Option 19 provides the user with the ability to set their own MAA weightings in the fitness function at the front, sides and rear. Finally, the slider highlighted as element 20 allows the user to trade-off between search speed and solution accuracy. If the user chooses speed over accuracy fewer angles are checked in the fitness function resulting in each solution being evaluated more quickly and vice versa.

4.10 Code testing

Before the design tool was used for deriving decoders the Tabu Search algorithm was tested to see if it was coded correctly and functioning as expected. The search was given the task of finding the optimum value of the Michalewicz test function (defined in chapter 2). The optimum value of this function is dependent on the number of parameters n so two different cases of varying levels of difficulty were chosen - $n = 2$ and $n = 5$ (in the latter the global minimum is more difficult to locate). In both cases 100 search runs were undertaken with the best solution found by the search recorded at the end of each run.

In the first case ($n = 2$) the search located the global minimum 94 times out of the 100 runs whereas in the second case ($n = 5$) the search located the global minimum 8 times out of the 100 runs. Please note that fewer optimum solutions were expected to be found for $n = 5$ because of the significant increase in the size of the search space. These results prove that the algorithm is correctly implemented because it is able to find solutions for a benchmark test problem. The number of times it locates the global minimum is comparable with other search algorithms.

In order to investigate whether the search was able to escape from a local minima the ‘current best’ solutions were recorded at each iteration of the search algorithm over a single run (for $n =$

2). Figure 4-11 plots the recorded values showing that after the search reaches the global minimum (at iteration 150) it selects a range of lower quality solutions in the hope of finding a better quality solution overall. This demonstrated that the Tabu list was working.

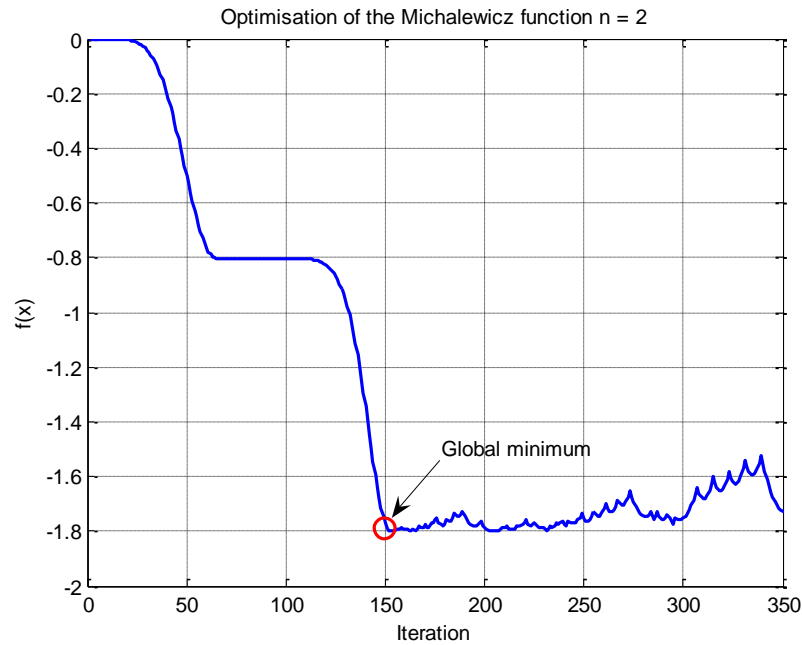


Figure 4-11: The ‘current-best’ solution recorded at each interaction of a single search run demonstrating the Tabu Search’s ability to escape from local minima.

The coding of the fitness function was tested by evaluating ‘good’ solutions in the performance panel of the design tool alongside solutions produced in other research work. Please note that the search would soon exploit any mistakes made in the coding of the fitness function which would easily be spotted when viewing the performance plots of generated solutions.

4.11 Summary

This chapter has given an overview of the decoder design tool developed in this research. Each component of the tool was identified and explained in detail. The next chapter will describe the detailed testing of each component.

Chapter 5

Theoretical Localisation Performance of the Developed Decoders

5.1 Introduction

This chapter examines the theoretical localisation performance of a range of ITU 5-speaker decoders derived using the design tool. The aim of this work was to assess the capabilities of each component of the system and provide the first steps towards validating the system as a whole. The derived decoders are analysed using the developed fitness function and the velocity and energy vectors to show their performance from a search optimisation point of view, and a decoder designer's point of view.

5.2 Design tool settings

Table 5.1 details the search settings that were applied in the options panel of the design tool when deriving the decoders in this chapter. The neighbourhood size and the Tabu tenure were set to increase with system order to take into account the greater number of decoder coefficients required per system order. As previously noted, twice the number of decoder coefficients for the neighbourhood size allows a positive and negative step to be made for each coefficient.

Bad moves	Neighbourhood size	Tabu tenure	Number of searches
250	2 x number of coefficients	2 x size of neighbourhood	10 runs of the design tool consisting of 100 searches

Table 5.1: Design tool search settings used when deriving the decoders

A fixed number of 1000 searches was chosen to allow a good range of solutions to be produced within a reasonable amount of time. The 1000 searches were divided into 10 runs of the design tool each consisting of 100 searches. A pair-wise comparison of the best solution from each run was undertaken with one selected as the best overall solution. In reality the best solution from a search run is the only solution the user would encounter when using the tool. Please note that this

configuration was used when deriving all decoders in this thesis unless explicitly stated otherwise.

All of the Ambisonic decoders presented in this chapter are frequency independent and were optimised for the ITU 5-speaker layout with rear speakers at $\pm 110^\circ$. The reader is reminded that a constant loudspeaker distance has been assumed for all decoders produced in this work.

5.3 Testing range-removal and importance

In chapter 4 the problem of objective dominance was discussed (see section 4.3). It was shown how the low frequency fitness function objectives (E_{LFFVol} , E_{LFFMag} , E_{LFFAng}) dominated the search for decoder coefficients because of their large range of potential values. In order to resolve this problem range-removal was included as a component of the design tool to ensure all objectives were constrained to the same range of values. A further concept termed ‘importance’ was added for logically biasing range-removed objectives.

In order to test range-removal the design tool was required to produce a first order frequency independent decoder. Two applications of the tool were undertaken: one with range-removal applied to the fitness function objectives and one without. In both applications no objective importance weightings were used.

Table 5.2 shows the objective values of the best solutions from both design tool applications. It can be seen that the design tool application without range-removal produced a solution dominated by the low frequency objectives. This is shown by the near ideal values for the low frequency objectives (E_{LFFVol} , E_{LFFMag} , E_{LFFAng}) for the best non range-removal solution. In contrast, the best solution derived using range-removal better meets all of the objectives simultaneously because all the objectives were treated equally in the search. For this decoder improvements were made for 4 out of 7 objectives (E_{HFFVol} , E_{HFFMag} , E_{HFFAng} , $E_{AngMatch}$) at the cost of the low frequency objectives. This demonstrates the effectiveness of using range-removal.

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
Range-removal <input type="checkbox"/>	0.0000	0.0352	0.2376	63.7042	0.1858	33.2987	33.3437
Range-removal <input checked="" type="checkbox"/>	0.0015	0.0346	51.5343	60.7148	3.7459	23.4983	21.9366

Table 5.2: Fitness function objective values of the best solutions encountered during design tool applications without the range-removal component and with the range-removal component.

The design tool performance plots for both decoders are shown in figure 5-1 and figure 5-2 respectively. Note that in figure 5-1 the velocity vector is ideal and the pressure (low frequency volume) is even around the listener. In figure 5-2 the velocity vector performance is reduced but the energy vector has been improved.

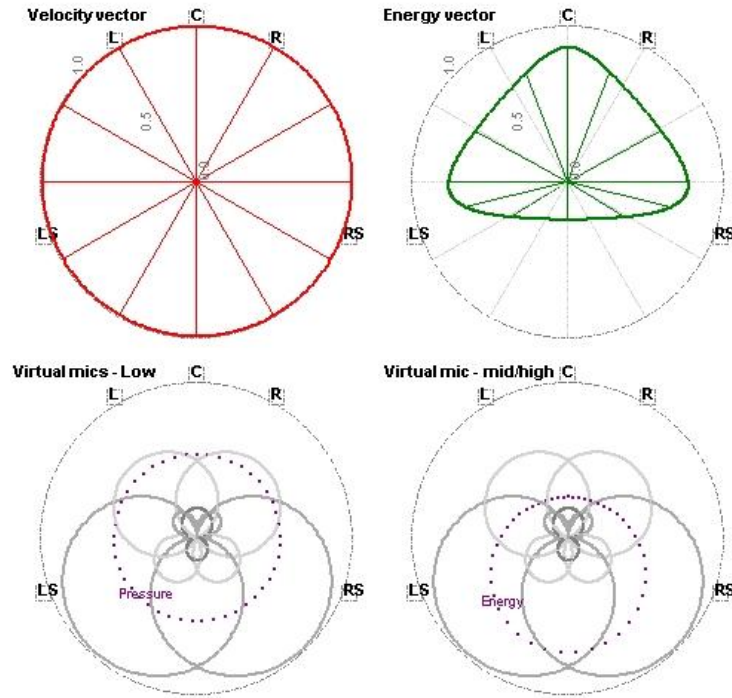


Figure 5-1: Performance plot of a first order decoder derived without range-removal.

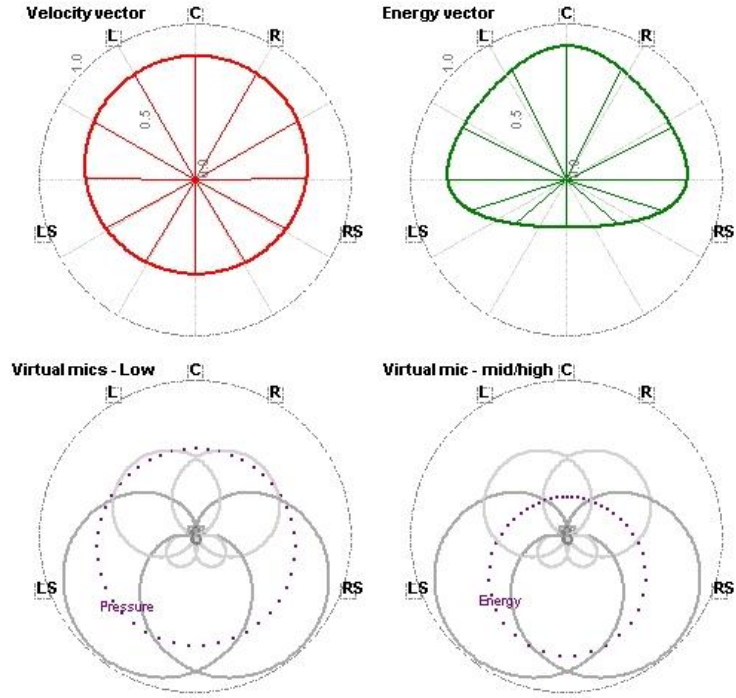


Figure 5-2: Performance plot of a first order decoder derived with range-removal.

Although range-removal resolves the problem of objective dominance, on its own it does not guarantee that an acceptable solution (from a decoder designer's point of view) will be produced by the search. Applying importance weightings to range-removed objectives allows a decoder designer to tailor performance towards specific desirable criteria. In order to demonstrate this, a further application of the design tool was undertaken with the aim of producing a decoder with improved mid/high frequency angle performance ($E_{HF\text{Ang}}$). The mid/high frequency angle objective was given an importance weighting of 10 while the other objectives had equal importance weights of 1. Table 5.3 shows the objective values for the best solution produced by for this application. The best equal importance solution is included for comparison.

	$E_{LF\text{Vol}}$	$E_{HF\text{Vol}}$	$E_{LF\text{Mag}}$	$E_{HF\text{Mag}}$	$E_{LF\text{Ang}}$	$E_{HF\text{Ang}}$	E_{AngMatch}
Importance weighted	0.1511	0.1192	85.7278	75.6519	6.7690	0.1825	6.7705
Equal importance	0.0015	0.0346	51.5343	60.7148	3.7459	23.4983	21.9366

Table 5.3: Fitness function objective values of the best solution produced by the design tool when giving higher importance to the mid/high frequency angle objective.

As expected, higher importance for the mid/high frequency angle objective led to an improvement for this objective when compared to the previously derived equal importance decoder. Please note, however, that selecting a higher weight for the mid/high frequency angle objective also led to improved performance for the angle match objective ($E_{AngMatch}$) and poorer performance for all other objectives. This shows that care needs to be taken when selecting importance weightings because of objective inter-dependency.

Figure 5-3 shows the performance plot for this decoder. Note the improved energy vector angle response when compared to the previously derived equal importance decoder displayed in figure 5-2.

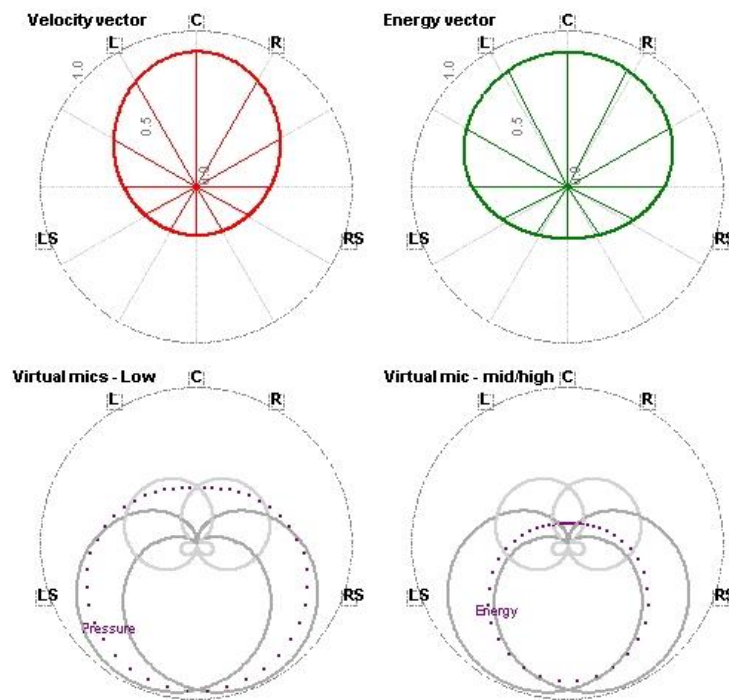


Figure 5-3: First order decoder derived with a greater importance given to the mid/high frequency angle objective.

Clearly, the main advantage of using range-removal in the design tool is objectives that are unrelated can be compared and evaluated together without the problem of objective dominance. When using range-removal together with importance, objectives that are deemed equally important should attain approximately the same level, in percentage terms, whereas more important objectives should be closer to their ideal values than less important objectives.

5.4 Evaluation of the improved multi-objective fitness function

Having shown the value of including range-removal and importance in the fitness function the next task was to directly evaluate the individual and combined impact of the new angle match objective ($E_{AngMatch}$) and the revised volume objectives (E_{LFVol} and E_{HFVol}). Five different applications of the design tool were undertaken:

1. In the first application the volume objectives and the angle match objective were switched off in the fitness function by applying importance weightings of 0.
2. In the second application the volume objectives from the work of Wiggins were switched on in the fitness function by applying an importance weighting of 1. The angle match objective was switched off.
3. In the third application the revised volume objectives replaced those by Wiggins and were switched on in the fitness function. The angle match objective was switched off.
4. In the fourth application the angle match objective was switched on but the revised volume objectives were switched off.
5. Finally, in the last application of the design tool, both the angle match objective and the revised volume objectives were switched on so their combined impact in the fitness function could be evaluated.

In all five cases the fitness function objectives were given equal importance (excluding the objectives under test).

Table 5.4 presents the objective values of the best solution found in each application. The best solution values presented in this table demonstrate that the new objectives are successful in meeting their goals. In the third application (row 3 of table 5.4) it can be seen that switching on the revised volume objectives resulted in the design tool producing a decoder with better volume performance when compared to the best solution produced in application one (row 1 of table 5.4) and application two (row 2 of table 5.4) (lower objective values are better).

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
Revised Volume <input type="checkbox"/> Angle <input type="checkbox"/>	0.1739	0.2226	55.9049	59.8127	0.0102	22.3771	22.3806
Wiggins volume <input checked="" type="checkbox"/> Angle <input type="checkbox"/>	0.0027	0.0890	37.4894	50.2268	0.0097	48.7330	48.7307
Revised volume <input checked="" type="checkbox"/> Angle <input type="checkbox"/>	0.0001	0.0416	56.8342	50.7238	5.7829	43.7601	42.5295
Revised volume <input type="checkbox"/> Angle <input checked="" type="checkbox"/>	0.5493	0.6276	84.0688	72.5466	0.0864	1.9488	1.9140
Revised volume <input checked="" type="checkbox"/> Angle <input checked="" type="checkbox"/>	0.0175	0.0170	55.3970	61.2816	3.3100	21.3271	19.7557

Table 5.4: Objective values for the best solutions produced when testing the impact of the new objectives added to the fitness function.

In the fourth application (row 4 of table 5.4) it can be seen that switching on the angle match objective (without the revised volume objectives) resulted in the design tool producing a decoder with velocity vector and energy vector angles that match much more closely than the best solution produced in application one. In this scenario it looks like there is a direct relationship between the angle match objective and the mid/high frequency angle objective (E_{HFAng}) because the low frequency angle objective gives similar performance while the mid/high frequency angle objective improves significantly. In application five switching on both the revised volume objectives and the angle match objective resulted in the design tool producing a solution that is better for volume and better for angle match, but not as good as when the two objectives are optimised individually.

In summary, this section has shown that the new angle match objective increases the possibility of deriving decoders with velocity vector and energy vector angles that match closely by angle

around the listener. According to Gerzon's definition of the Ambisonic system this is a desired performance characteristic and has been neglected in previous work (Gerzon & Barton 1992). In addition, the revised volume objectives are able to generate decoders with even volume performance around the listener, also better meeting Gerzon's criteria.

5.5 The generation of higher order decoders

The next task involved assessing the design tool's capability of deriving higher order decoders. The aim was to produce second order, third order and fourth order frequency independent decoders. Equal importance weightings were used during the search.

Table 5.5 presents the objective values for the best solutions produced for each order. The best first order decoder derived in previous section is shown for comparison. In this table the total fitness values highlight the performance transition that can be achieved when increasing the decoder order - as the decoder order increases the total fitness values of the solutions improve.

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$	<i>Total</i>
1st order	0.0175	0.0170	55.3970	61.2816	3.3100	21.3271	19.7557	161.1059
2nd order	0.0063	0.0063	44.0816	32.7134	16.1490	35.8677	23.2759	152.0978
3rd order	0.0167	0.0020	32.3641	32.6588	10.1408	38.3335	28.7547	142.2706
4th order	0.0031	0.0044	41.1994	46.9113	13.0542	15.2001	25.0529	141.5589

Table 5.5: Best solutions produced for each decoder order in an equal importance application.

The better performance for the higher order decoders is due to the fact that higher order decoders produce virtual microphone responses which are more fitting to the 5-speaker layout. In order to illustrate this, figure 5-4 shows typical virtual microphone responses for optimised ITU 5-speaker decoders from first order to fourth order. Clearly for the fourth order decoder the responses are much more directional at the front of the system where the loudspeakers are closer together. Note also as the decoder order increases the more the centre loudspeaker is used.

At the rear the differences between the virtual microphone responses with system order are a bit more subtle. As the decoder order increases the virtual microphones become asymmetrical although remain quite similar to the first order responses.

The best fourth order decoder derived in this work was selected for further tests (described in the following chapters). Figure 5-5 displays the performance plot for this decoder. When comparing this decoder with the best first order decoder from the end of section 5.4 it can be seen that much better vector magnitudes are produced, particularly around at the front of the system.

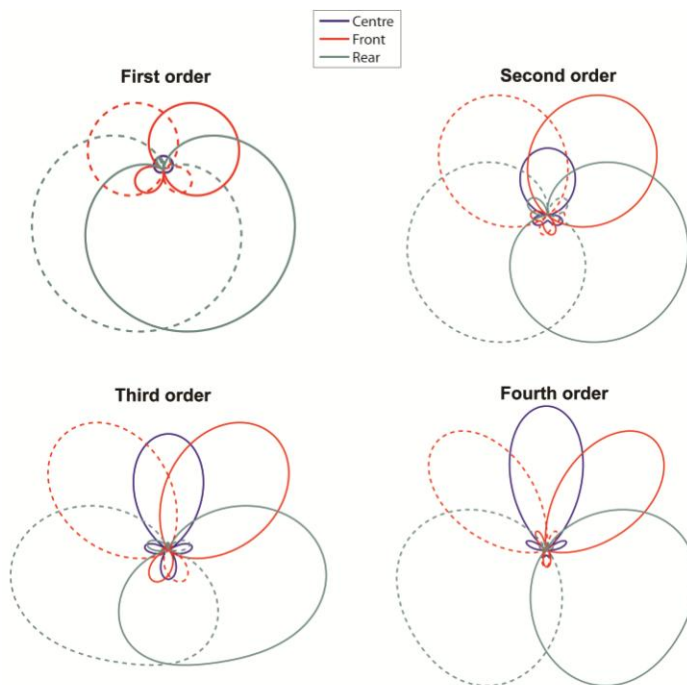


Figure 5-4: Virtual microphone response of typical decoders from 1st order to 4th order

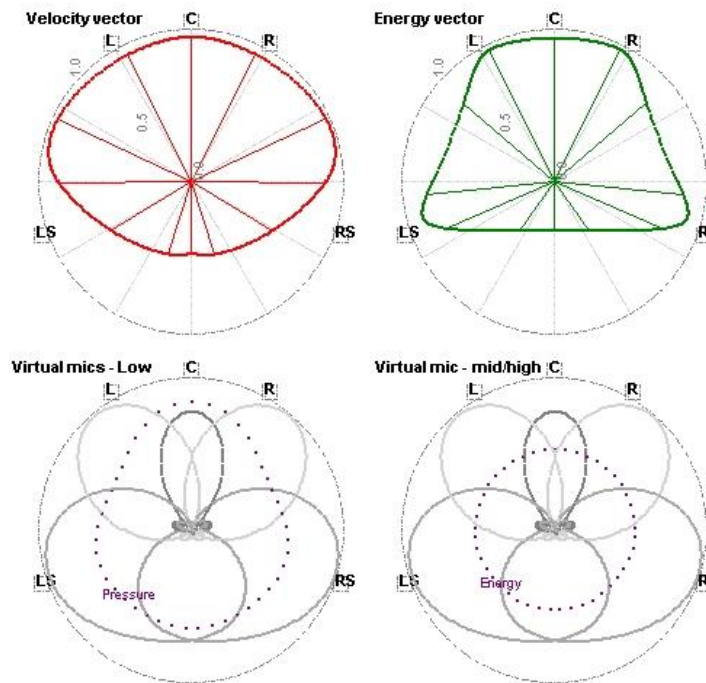


Figure 5-5: A good fourth order decoder derived using the design tool

5.6 Evaluation of the even performance optimisation component

The next task presented to the design tool was to derive a decoder with even localisation performance by angle around the listener. The aim was to investigate the capability of the even error optimisation component of the design tool when used in combination with range-removal and importance. The desired decoder was a fourth order frequency independent decoder.

Two applications of the design tool were undertaken to test the even error component. In the first application each of the fitness function objectives (including the even error objectives) were given equal importance weightings of 1. In the second application the importance of the even error objectives was increased to 2. In the following analysis the best decoders produced in each application are referred to a Decoder A and Decoder B.

Figure 5-6 shows the total performance error by angle for Decoder A and Decoder B (summed objective error by angle). For comparison the response of the standard fourth order decoder derived in section 5.5 is included. The mean of the total error is provided in each plot for reference.

In terms of overall localisation performance Decoder A is quite similar to the standard fourth order decoder. However, the localisation performance of Decoder A is more even at the front and the sides of the system (between 0° and 120°). Decoder B has the most even total performance error distribution of all three decoders reflecting the performance weightings that were used. However, the increase in even performance has been at the cost of a reduction in overall performance (note the higher error value).

Figure 5-7 plots the individual objective values by angle for each decoder (the volume objectives are omitted from this analysis as they were originally designed to ensure even error). It is clear that Decoder B has the most even performance for all objectives when compared to the standard fourth-order decoder and Decoder A. This is confirmed in table 5.6 which gives the standard deviation for all objectives for each of the decoders. When compared to the standard fourth order decoder, Decoder A has more even performance for the low frequency angle objective (E_{LFAng}), the low frequency magnitude objective (E_{LFMag}) and the high frequency magnitude objective (E_{HFMag}).

Total performance error by angle

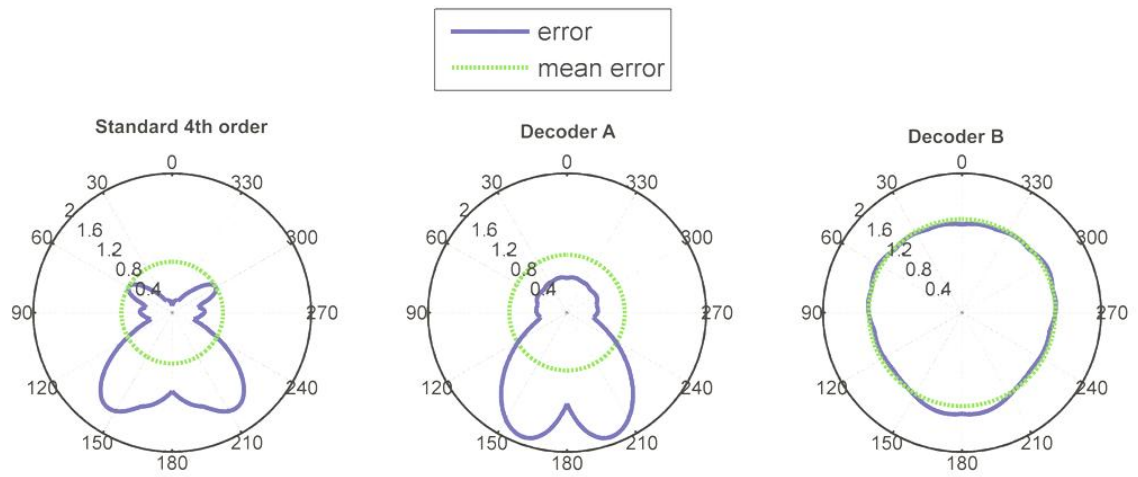


Figure 5-6: Total performance error by angle for the even error optimised decoders and a typical decoder. The standard deviation and mean of the error are included for comparison.

Objective error by angle

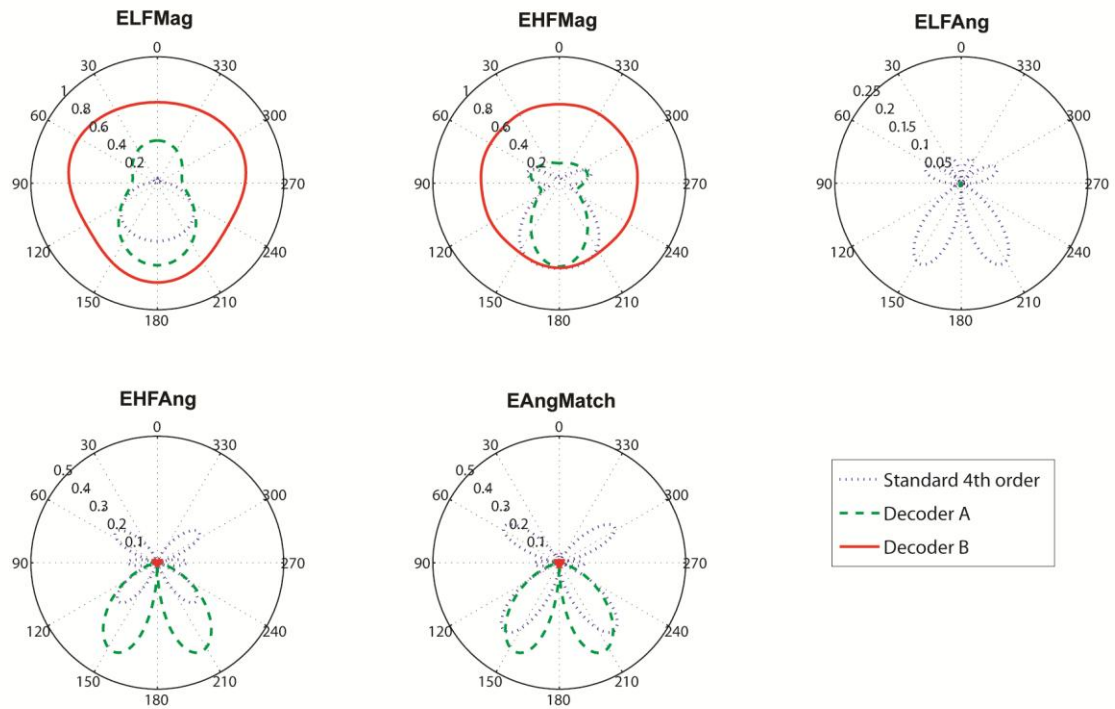


Figure 5-7: Objective error by angle for all three decoders (note the change in scale in each plot)

Objective	Typical	Decoder A	Decoder B
E_{LFAng}	0.0516	0.0004	0.0019
E_{HFAng}	0.0715	0.1383	0.0051
$E_{AngMatch}$	0.1057	0.1382	0.0055
E_{LFMag}	0.1777	0.1426	0.0389
E_{HFMag}	0.2201	0.1499	0.0127
E_{LFVol}	0.0002	0.0002	0.0001
E_{HFVol}	0.0000	0.0000	0.0000

Table 5.6: Standard deviation of objective error for all three decoders

Whilst searching for even error decoders an interesting objective inter-relationship became apparent. When a low error value was obtained for the vector angle objectives, a high error value was obtained for the vector magnitude objectives and vice versa (see Decoder B's performance in figure 5-7). Decoder designers using even error design criteria in future work should take this inter-dependency into consideration when selecting importance weightings.

In summary, this analysis demonstrates the use of the even error optimisation component incorporated into the design tool. The results show that the even error objectives are able to reduce significantly the large variation in performance around the 360° sound stage. However, consideration should be made when determining their importance weightings. It was found that there is a direct tradeoff between choosing good overall performance and good even performance by angle for each of the objectives. However, by adjusting the importance weighting between the original improved fitness function objectives and the even error objectives, a decoder designer can achieve the required balance between good overall decoder performance and even performance for all angles.

Following this work a further design tool application was undertaken with the aim of producing an even error decoder for further evaluation in later experiments (described in the chapters 6 and 7). After several search runs with different importance weights a decoder was found with suitable characteristics. Table 5.7 details the importance weights that were used.

E_{LFBol}	E_{HFBol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
0.0	1.0	1.0	1.8	1.0	4.0	1.0
$E_{LFAngEv}$	$E_{HFAngEv}$	$E_{LFMagEv}$	$E_{HFMagEv}$			
1.0	1.0	1.0	1.0			

Table 5.7: Even error decoder objective importance weightings

Note that in this application the low frequency volume objective was effectively switched. The reason for this is because the energy is more suited to represent the perceived volume for the listener for a frequency independent decoder (Gerzon & Barton 1998). In addition higher importance weightings were given to the energy vector magnitude objective (E_{HFMag}) and the energy vector angle objective (E_{HFAng}) (i.e. aka a max r_E decoder).

Even error decoder - design tool performance plot

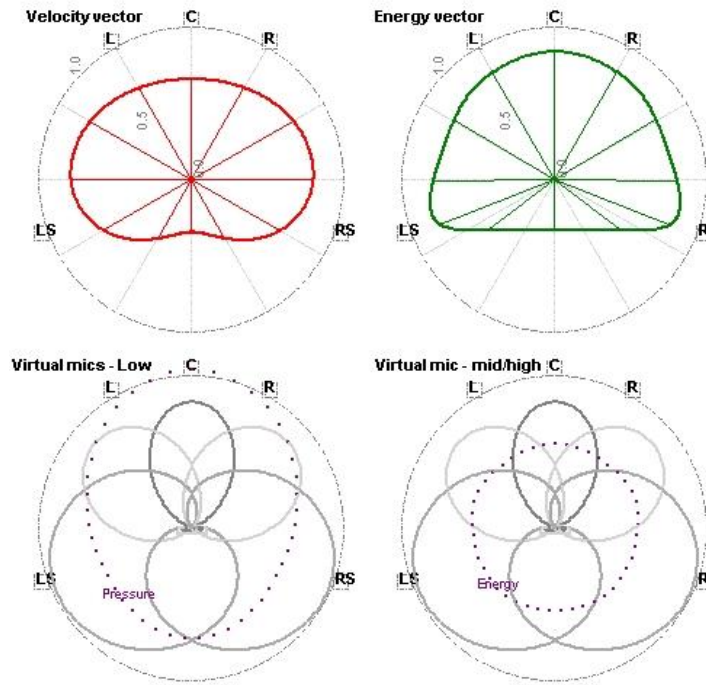


Figure 5-8: The decoder design tool performance plot for the even error optimised decoder

The derived decoder has fairly even performance by angle without reducing the overall performance (see figure 5-8). The energy vector and velocity vector responses are comparable

apart from at the front and rear of the system where the energy vector is better (as desired). Also, note the constant energy level around the listener.

5.7 Evaluation of the minimum audible angle optimisation component

The next design application that was presented to the tool was to produce a fourth order frequency-independent decoder with improved performance in directions where humans are more sensitive to sound localisation. The aim of this application was to investigate the capabilities of the MAA weighting component incorporated into the design tool. Equal importance weightings were given to all fitness function objectives when deriving this decoder.

Figure 5-9 shows the performance plot of the best decoder derived with the MAA component turned on.

With MAA component - design tool performance plot

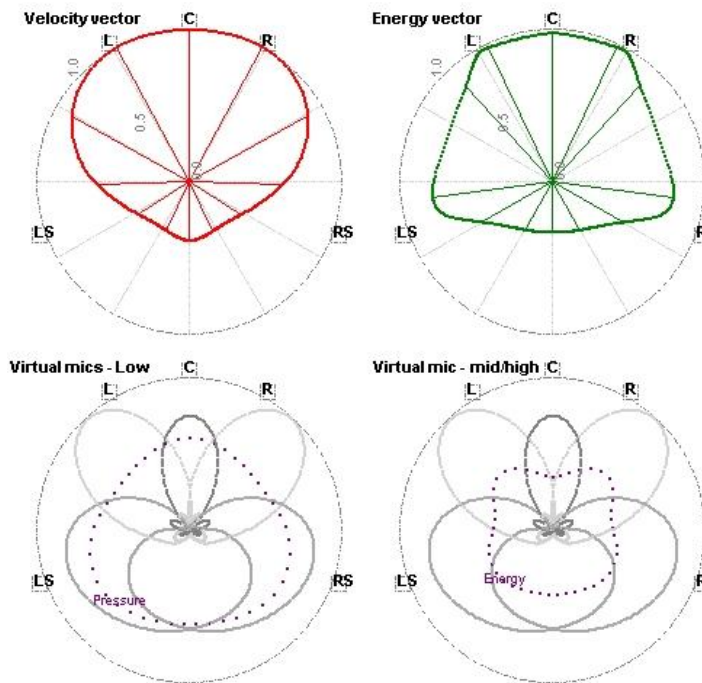


Figure 5-9: The decoder design tool performance plot for the fourth order MAA optimised decoder

It is clear that this decoder has much better performance at the front of the system when compared to the previously derived fourth order decoders in this chapter (see figure 5-2 and figure 5-8 for example). The vector magnitudes are very close to their ideal value of 1 in the front of the system. This increase in performance at the front has been at the cost of reduced localisation performance to the sides following the pattern of human spatial resolution (note the reduced performance of the energy vector angle in particular).

Although there has been a slight performance increase at the direct rear for the energy vector magnitude it is hard to improve the energy vector magnitude in this area because of the large angular spacing between the rear loudspeakers. In fact, the energy vector magnitude would theoretically only be able to reach a maximum value of 0.34 at 180° when the loudspeakers are arranged in this way (equivalent to pair-wise constant power panning) (Craven 2003). Furthermore, if this theoretical maximum was reached it is likely to have an adverse affect on other elements of a decoder's performance because of objective inter-dependency. One way of improving the theoretical localisation performance at the rear of the system is to reduce the angular spacing between the rear loudspeakers (David Moore & J. P. Wakefield 2008).

In summary, this work has demonstrated that by using the MAA component of the design tool it is possible to produce decoders with improved theoretical performance in directions where humans are more sensitive to sound localisation. The fourth order MAA decoder analysed in this section was selected for the experiments presented in the following two chapters.

5.8 Evaluation of the off-centre optimisation components

The next element of the design tool to be tested was the off-centre optimisation component. The goal was to produce a fourth order frequency-independent decoder with improved performance in off-centre listening positions. When deriving the decoder, 9 evenly distributed listening positions were evaluated in the fitness function: the centre point and 8 surrounding positions (see figure 5-10). Equal importance weightings were given to all objectives.

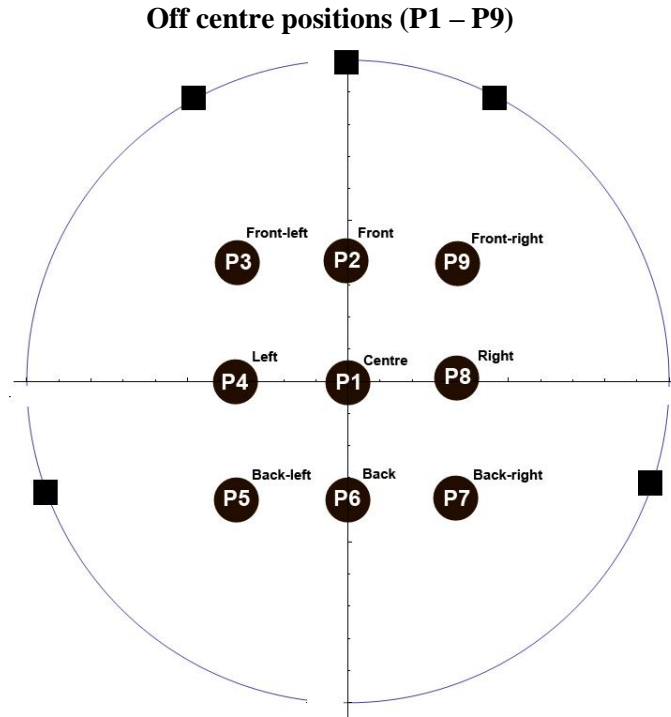


Figure 5-10: The off-centre positions that were evaluated in the improved fitness function.

Positions 2, 4, 6 and 8 are at 35% of the loudspeaker rig radius whereas positions 3, 5, 7 and 9 are at 50% of the loudspeaker rig radius. These positions were specifically chosen so in later practical experiments the same positions could be evaluated by listeners.

It is important to note that there is a direct performance trade-off when using this particular off-centre optimisation strategy. Improving the velocity vector or energy vector at one position can have an adverse effect on performance at another position because of the change in loudspeaker level.

Figure 5-11 and figure 5-12 plots the local velocity vector for the fourth order off-centre optimised decoder at the 9 listening points evaluated in the improved fitness function. The vectors are shown at 0°, 30°, 60° and 90° in figure 5-11 while the vectors are shown at 120°, 150° and 180° in figure 5-12. In each plot the local velocity vectors for the standard fourth order

decoder (from section 5.5) and the first order decoder (from section 5.3) are shown for comparison. An ideal vector is also indicated at each position.

The velocity vector performance of the off-centre optimised decoder is better at most positions and for most angles. Take, for example, when a source is panned to 120° . The local velocity vectors are closer to the ideal vectors (in terms of magnitude and angle) in nearly all listening positions.

Figure 5-13 and figure 5-14 shows the local energy vectors for the decoders. The difference in performance is again clear. For instance, when a source is panned to the front (0°) the local energy vector is closer to the ideal vector at all positions for the off-centre optimised decoder. For the other decoders, the vector angles are biased towards the front left loudspeaker when evaluated from positions 3, 4 and 5, and the front right loudspeaker when evaluating from positions 7, 8 and 9.

The most problematic area of the sound stage for all decoders is at the rear (see 150° and 180°). In off-centre positions the local velocity vectors and energy vector pull away from their ideal direction towards the nearest loudspeaker. This result was expected considering the large angular spacing between the rear loudspeakers.

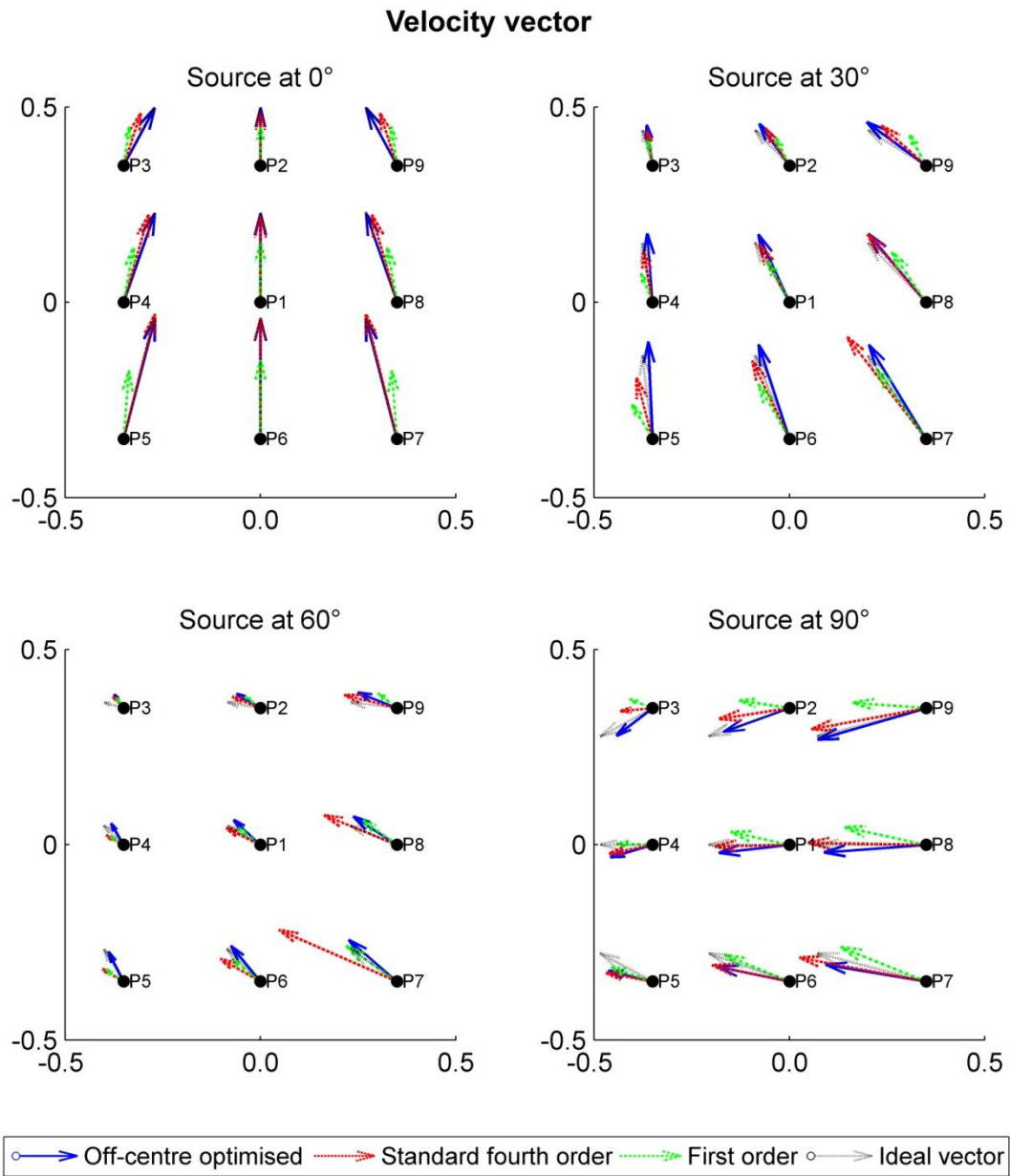


Figure 5-11: Local velocity vectors at each position evaluated in the improved fitness function for the angles 0° , 30° , 60° and 90° . Note the vector magnitudes at each position have been scaled to allow for better viewing.

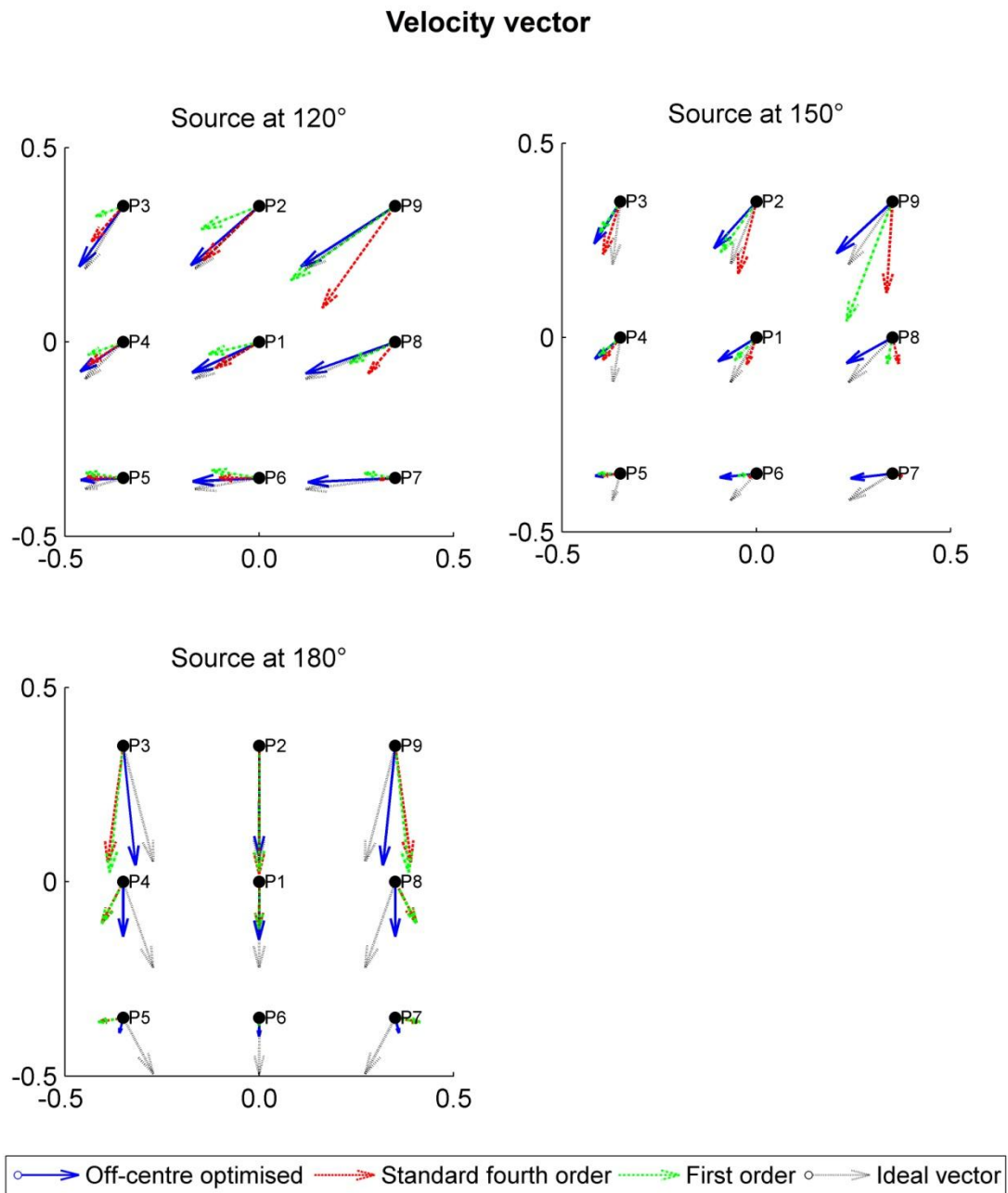


Figure 5-12: Local velocity vectors at each position evaluated in the improved fitness function for the angles 120°, 150° and 180°. Note the vector magnitudes at each position have been scaled to allow for better viewing.

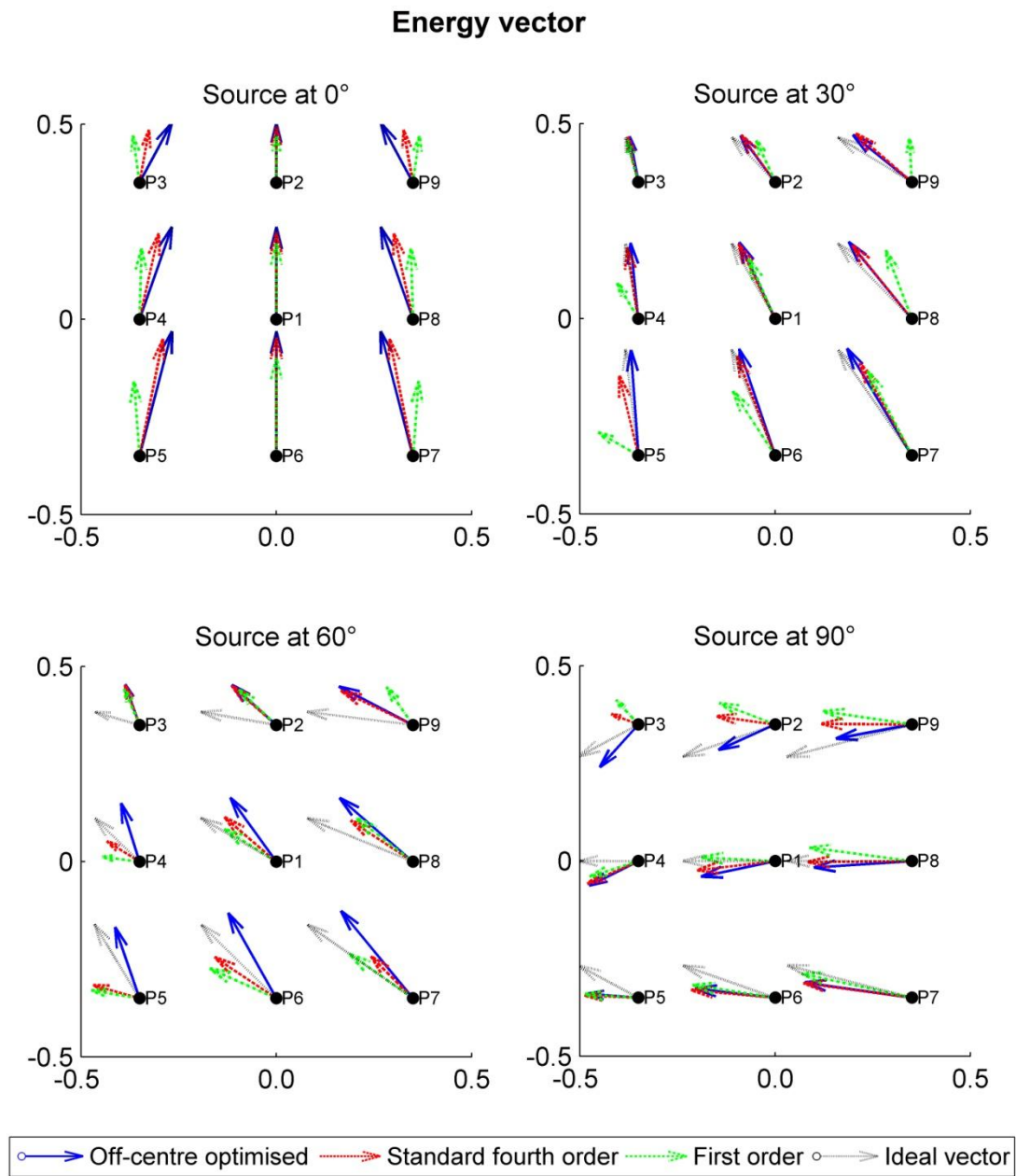


Figure 5-13: Local energy vectors at each position evaluated in the improved fitness function for the angles 0° , 30° , 60° and 90° . Note the vector magnitudes at each position have been scaled to allow for better viewing.

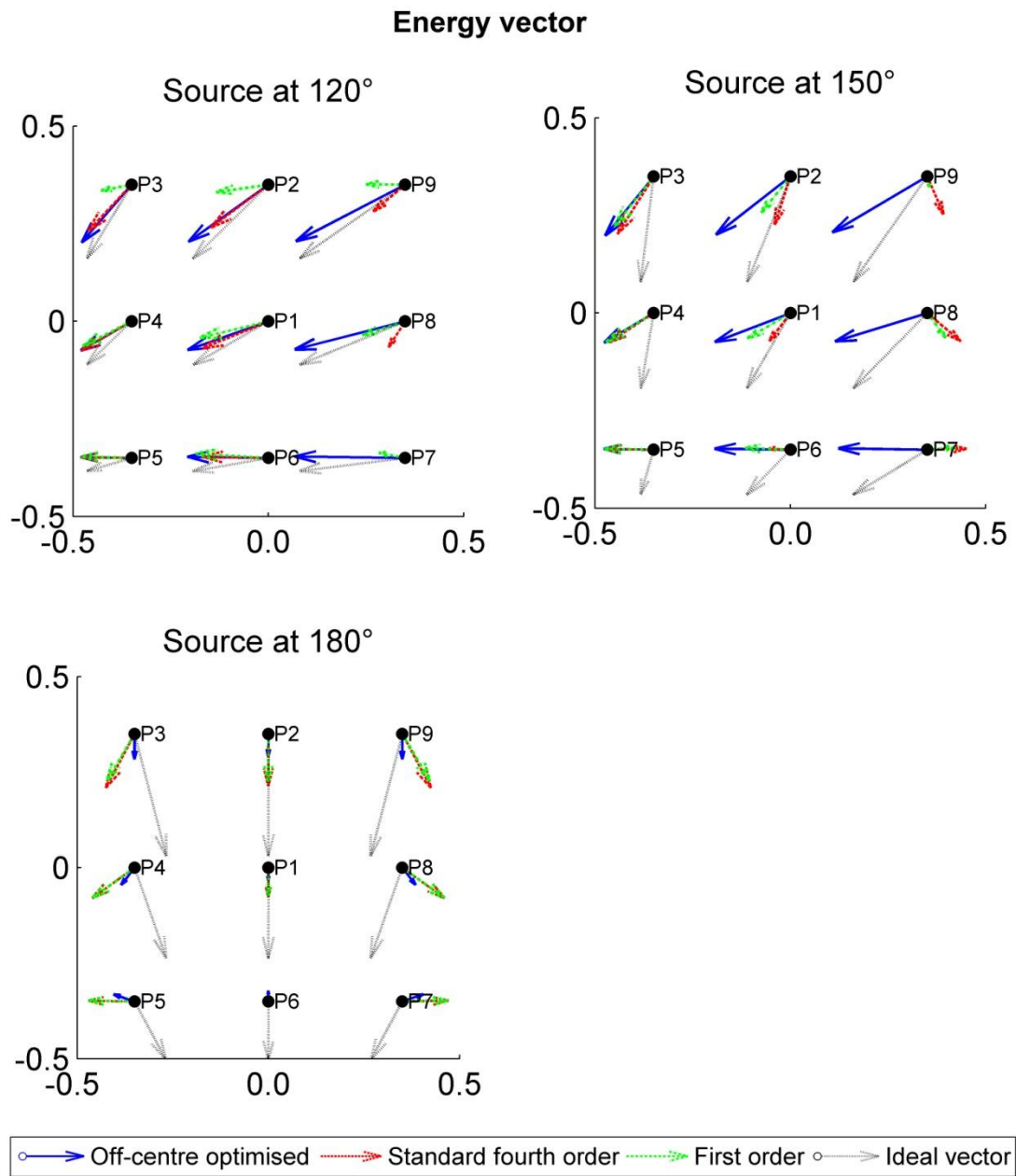


Figure 5-14: Local energy vectors at each position evaluated in the improved fitness function for the angles 120°, 150° and 180°. Note the vector magnitudes at each position have been scaled to allow for better viewing.

In order to investigate further the performance of the off-centre decoder, figure 5-15 plots the mean error of the velocity vector and energy vector magnitude and angle for each position taking into account each different source angle checked in the fitness function (i.e. one degree steps between the front and the rear). This figure demonstrates that the off-centre optimised decoder is able to produce better performance at a greater number of positions than the other decoders. Of particular note are the consistency low vector magnitude errors across all positions and the improved vector angles at listening position on the left side of the system and the right side of the system.

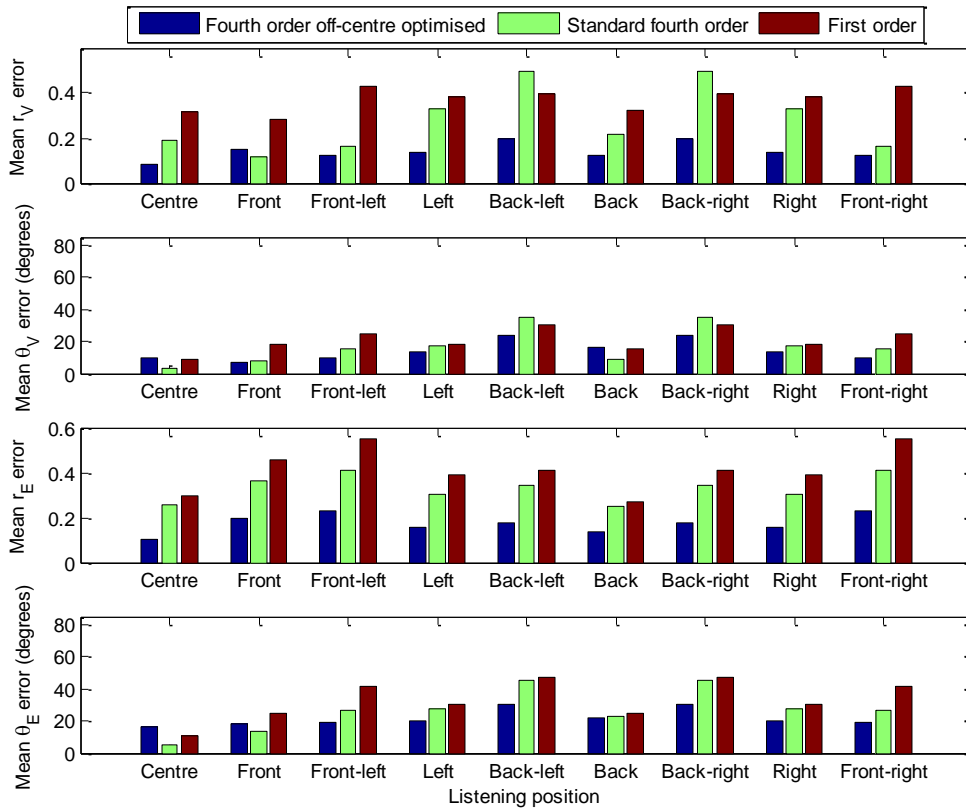


Figure 5-15: Mean magnitude and angle error for the velocity and energy vectors at each position.

As highlighted earlier, this off-centre optimised decoder was derived using equal importance. By using different weightings the decoder designer could improve the performance of the vector angles or magnitudes further.

In summary, this section has shown that the design tool is able to produce decoders with improved theoretical localisation performance in off-centre positions. A further application of the design tool was undertaken with the aim of producing an off-centre optimised decoder to be evaluated in further experiments. When deriving the decoder the following objective importance weightings were used during the search (see table 5.8). Greater importance was given to the mid/high frequency objectives ($E_{HF\text{Ang}}$, $E_{HF\text{Mag}}$, $E_{HF\text{Vol}}$, $E_{HF\text{AngEven}}$, $E_{HF\text{MagEven}}$) because the energy vector is believed to be a better predictor of sound localisation when in off-centre listening positions (Gerzon 1974; Gerzon & Barton 1992; Gerzon 1992b; Gerzon 1992a).

	$E_{LF\text{Vol}}$	$E_{HF\text{Vol}}$	$E_{LF\text{Mag}}$	$E_{HF\text{Mag}}$	$E_{LF\text{Ang}}$	$E_{HF\text{Ang}}$	E_{AngMatch}
Importance weights	0.0	0.9	0.25	1.0	0.25	1.0	0.5
	$E_{LF\text{AngEv}}$	$E_{HF\text{AngEv}}$	$E_{LF\text{MagEv}}$	$E_{HF\text{MagEv}}$			
Importance weights	0.25	1.0	0.25	1.0			

Table 5.8: Objective importance weightings used when deriving the off-centre optimised decoder

5.9 Search algorithm acceleration using High Performance Computing hardware

The final component of the design tool to be tested was the ability to run searches on High Performance Computing (HPC) hardware. When the user switches this component on, all searches automatically run remotely on a computer server equipped with 2 ClearSpeed x620 Application Accelerator boards. One search run using the HPC hardware consists of 1536 search instances executing in parallel (i.e. 4 chips each with 96 processor elements, 4 calculations per processor element). After the remote searches finish the results are transferred to the user's computer and can be displayed using the design tool.

In order to investigate the speed increase the user could gain from using the HPC component an average time was taken from 10 search runs. In this investigation three versions of the search were executed:

1. **HPC version** – run remotely on a server equipped with the ClearSpeed boards. Each search instance is started from a random point and stops after a fixed number of 1000 moves.
2. **Reference version** - run on a modern day computer. This version of the search is identical to the HPC version except a search run consists of running 1536 searches in sequence rather than in parallel.
3. **Standard design tool version** – run on a modern day computer. This version of the search is used when the HPC component is switched off on the design tool. In contrast to the other versions a Tabu list is used and the search is stopped after a fixed number of *bad* moves. Because of these additions we expect it to take longer to run. However, it could potentially yield a better quality solution. A search run for this version consisted of evaluating 1536 searches in sequence rather than in parallel.

In all three versions the aim was to produce a first order frequency independent decoder. Equal importance weightings were given to all fitness function objectives.

Table 5.9 shows the average times for each version taken from the 10 search runs. Also shown are the total fitness values for the best solutions produced from each version as well as the total fitness.

	Average search run time	Best fitness
Standard	54mins 51secs	160.8412
Reference	22mins 30secs	161.4635
HPC	54secs	161.9850

Table 5.9: Comparison of the different search versions.

The results show that the search on the HPC hardware was approximately 24 times faster than the reference and approximately 59 times faster than the standard design tool version. Over the 10 search runs the standard version was able to find the best solution. This result was expected given that the standard version of the search employed a Tabu List to allow the search to escape from local minima and was stopped after a fixed number of bad moves. However, the difference between the best solution produced by the standard version and the best solution produced by the HPC version is only very slight. Figure 5-16 shows the mean total fitness and 95% confidence intervals for each version given the best solutions from the 10 search runs.

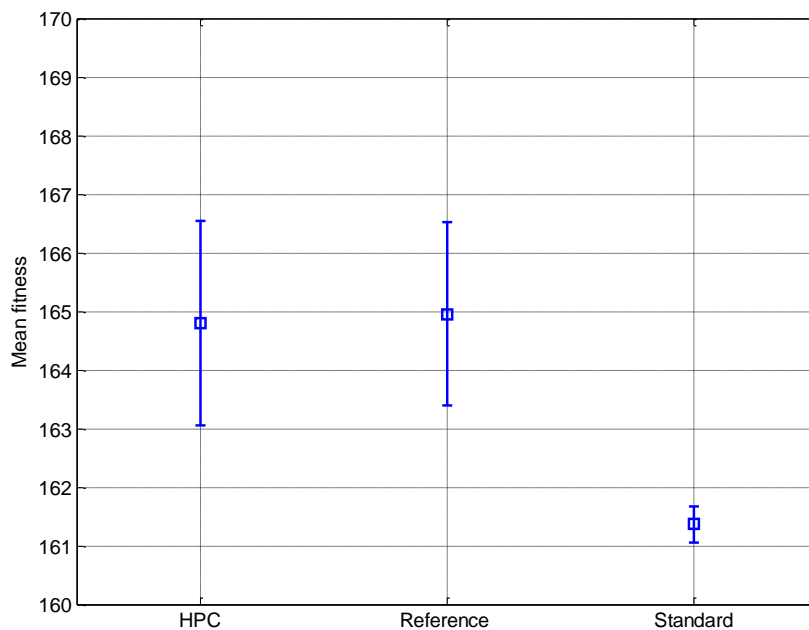


Figure 5-16: Mean total fitness and 95% confidence intervals

In summary, the HPC component of the design tool improves the speed of the search process. This speed increase is important as it allows more solutions to be evaluated within a set amount of time increasing the likelihood of finding a good solution. Furthermore, from a decoder designer's point of view, this speed increase means the tool can be much more interactive – it allows almost real time experimentation with different decoder design criteria.

The benefit of using the standard search is the user can run it locally on their computer and potentially find a better solution if time is not an issue.

5.10 Summary

This chapter investigated the capability of each component of the design tool. Firstly, in section 5.3, range-removal was shown to resolve the problem of objective dominance allowing solutions to be derived that better meet all of the fitness function objectives simultaneously. Using range-removal in conjunction with importance allows a decoder designer to systematically and logically bias the search in favour of specific fitness function objectives giving more flexibility when tailoring the performance of a decoder.

Next, in section 5.4, the multi-objective fitness function was evaluated. The new angle match objective ensures decoders derived using the design tool more closely match Gerzon's requirements for the Ambisonic system. The volume objectives are able to produce decoders with close to ideal volume performance when included in the fitness function.

In section 5.5 the even error optimisation component was evaluated. The design tool was used to produce fourth order decoders with even localisation performance by angle around the listener. The produced decoders demonstrate the effectiveness of the even error component. It was shown that the extent of the even performance can be controlled using importance weighting.

Section 5.6 demonstrated the MAA component is able to improve the theoretical localisation performance of decoders in directions where humans are more sensitive to sound localisation. A fourth order decoder was derived with near ideal performance at the front of the system at the cost of performance to the sides. Performance for this decoder was also improved in the problematic area at the rear of the system.

Section 5.7 showed that the off-centre optimisation component is able to improve the theoretical localisation performance of decoders in distributed listening positions.

Finally, section 5.8 showed how the time-to-solution can be reduced by using HPC hardware. Shorter search times improve the tool's level of interactivity.

Selected decoders produced during this work were further assessed by listening tests with human subjects (presented in the next chapter) and binaural measurements (presented in chapter 7).

Chapter 6

Psychophysical Evaluation of the Developed Decoders

6.1 Introduction

This chapter describes a series of listening tests designed to further assess the localisation performance of decoders produced using the design tool. Listening tests were particularly important for investigating how the human auditory system interprets the effects of the different optimisation methods incorporated in the design tool. The overall aim was to validate each of the design tool's components by producing a good decoder according to what the components aim to achieve.

The following series of tests was performed:

1. Localisation of real sound sources
2. Localisation of decoded sound sources from the central listening position
3. Localisation of decoded sound sources from distributed listening positions

In the first test, listeners were assessed on their capability of localising real sound sources positioned at discrete angles in the horizontal plane. The aim of this test was to produce a set of results that define a 'best case' for localisation accuracy. In the second test, listeners were required to localise panned sound sources from the central listening position to investigate the performance of decoders optimised for the central listening position. In the third and final test, listeners were required to localise panned sound sources from distributed listening positions to investigate the performance of the off-centre optimised decoders.

In this experiment the assumption was made that human localisation is equally capable on the left and right sides. This means that localisation need only be assessed on one side of the listener which can be used to reduce the number of evaluations the listener needs to make. This approach

was used so the decoders could be tested to a greater angular resolution without risking listener tiredness which could potentially influence the results. In support of this, much of the empirical data in the literature shows the capability of our hearing system is approximately symmetrical. For example, in their extensive study of sound localisation Oldfield and Parker found no differences in localisation accuracy between the left and right sides (Simon R Oldfield & Parker 1984). See also the extensive study of the human auditory system detailed by Blauert (Blauert 2001) and the work by Makous and Middlebrooks (Makous & Middlebrooks 1989).

6.2 Experimental setup

The experiment was conducted in a music studio at the University of Huddersfield. The dimensions of the room at floor level are 4.5m (L) x 5.5m (W) x 2m (H). The reverberation time of the room (RT 60) is detailed in figure 6-1. The broadband ambient noise level of the room is approximately 21dB (A).

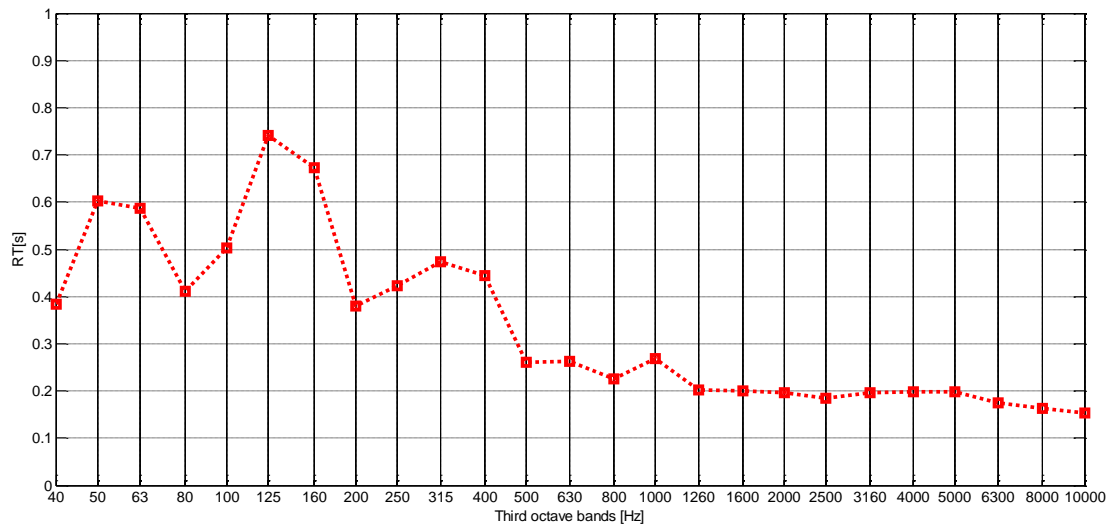


Figure 6-1: RT60 of the music studio used for the listening tests

The loudspeaker array used in the tests consisted of Genelec 8040A loudspeakers. In the localisation of real sources, 19 Genelecs were arranged every 10 degrees around the listener from 0 degrees to 180 degrees in a semi-circle with a radius of 2m to the right of the central listening position. Each loudspeaker was clearly labeled so the subjects could identify its location in

degrees. In the decoded source tests 5 Genelecs were arranged according to the ITU 5-speaker specification (rear speakers at $\pm 110^\circ$). All loudspeakers were equidistant from the centre point and were more than 0.5m away from the nearest wall. Figure 6-2 shows the geometry of the loudspeaker array in the room.

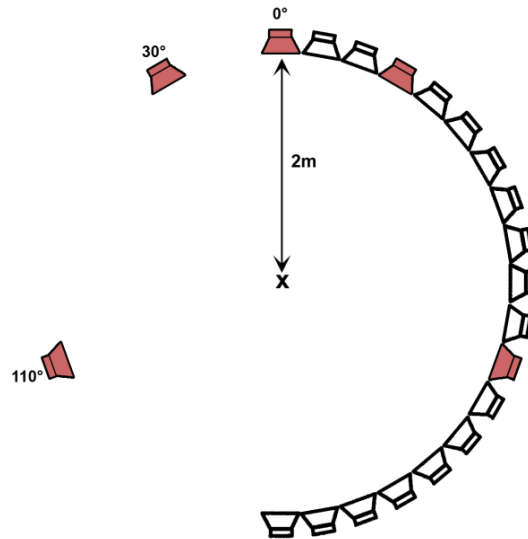


Figure 6-2: Geometry of the loudspeaker array in the listening tests

All loudspeakers were calibrated to a sound pressure level of 85dB(A) at a distance of 30cm on-axis with the tweeter. The sound pressure level at the central listening point was ≈ 70 dB(A).

Sound source localisation is very much dependent on the nature of the sound (see chapter 2). The frequency content of the source signal and the amplitude envelope of the source signal play an important role (Brian Moore 2003). In the light of this, three different source signals were employed in the tests: low frequency noise bursts (< 700 Hz), mid/high frequency noise bursts (700 Hz – 5000 Hz) and continuous male speech. The band-limited noise bursts were specifically chosen so that localisation performance could be tested in the low and mid/high frequency regions of the human hearing range (i.e. the frequency ranges that the velocity vector and energy vector broadly correlate with). Figure 6-3 shows how the noise bursts were presented to the listener.

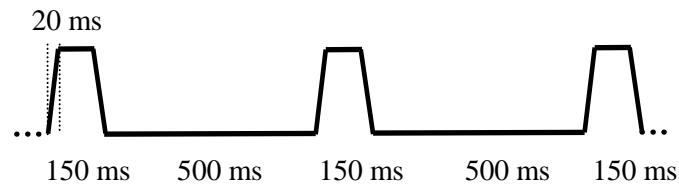


Figure 6-3: The amplitude envelope of the low and mid/high frequency noise stimuli

The bursts had a length of 150ms and were repeated three times with a break of 500ms between each burst (total length of 1450ms). The attack and release time of the amplitude envelopes was 20ms. The short burst times were specifically chosen to limit the likelihood of the listeners using head turning cues during the tests. Research by Makous and Middlebrook demonstrated that similar stimuli parameters were favourable in this respect (Makous & Middlebrooks 1989).

The continuous male speech was chosen because it is a source that has been shown to be easy to localise in many similar tests (Blauert 2001; Bates, Kearney, Boland et al. 2007). It contrasts with the noise bursts and represents a more “real world” signal that a surround sound system might typically be used to reproduce. The length of the speech signal was approximately 5 seconds and was only played once to the listeners. All source files had a resolution of 16 bit and a sampling rate of 48 kHz.

6.3 Test subjects

A total of 14 subjects took part in the tests (11 male and 3 female). All subjects were within the age range of 20-45 (average age of 25) and had no known hearing impairments. Most had not taken part in a formal listening test before but were accustomed to surround sound listening through personal equipment or through academic study. To ensure all subjects were of a similar standard a short training session was given before the start of each test. They were also given the opportunity to take a mock test where the results were given as feedback to allow them to stabilise their performance.

6.4 Test 1 - Real sound source localisation

6.4.1 Test procedure

In this test subjects were presented with one stimulus at a time from a randomly chosen loudspeaker at 10 degree intervals between 0 and 180 degrees. Their task was to correctly identify the loudspeaker that emitted the sound. This procedure was repeated until every loudspeaker in the array had played each of the 3 different stimuli once (57 sounds in total).

A software application was created which the listener operated during the test on a laptop computer (figure 6-4 shows the user interface). The software was designed to be simple to operate to avoid the user being distracted from the task in hand. For example, each of the user interface elements could only be selected in a specific order to prevent listeners from accidentally selecting the wrong option. In addition, an on-screen countdown from 3 seconds was given after the user clicked the “play sound” button so users could ready themselves before a sound was presented. Subjects were played the sound once and had to select an angle before proceeding further (forced-choice).

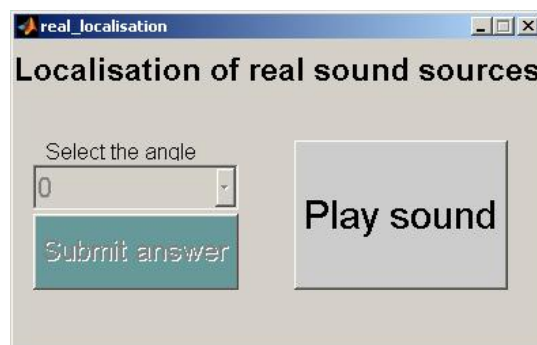


Figure 6-4: User interface of the real source listening test software

The subjects were positioned at the central listening point (2 metres from the loudspeakers and at the same height at ear level). Before the test started the test instructor ensured the subjects' heads were aligned with the loudspeakers at 0 degrees (in front) and 90 degrees (to the side). During the test each subject was asked to keep their head as still as possible while the sound was playing

to limit the influence of head-turning cues. They were instructed that they could move their heads after the sound had finished playing to confirm the angle of the loudspeaker. The test took, on average, 15 minutes to complete.

6.4.2 Results

Figure 6-5, figure 6-6, and figure 6-7 present the results for the respective sound sources. In each figure the subject's response is plotted with respect to the actual source angle. In all three tests front-back localisation confusions are apparent (see all responses within or on the border line of the shaded regions)³. As was highlighted in the chapter 2, front-back confusions generally occur when head turning cues are not available to the listener, and also when the interaural time difference and interaural level difference are the same in the front and the rear (i.e. the cone of confusion).

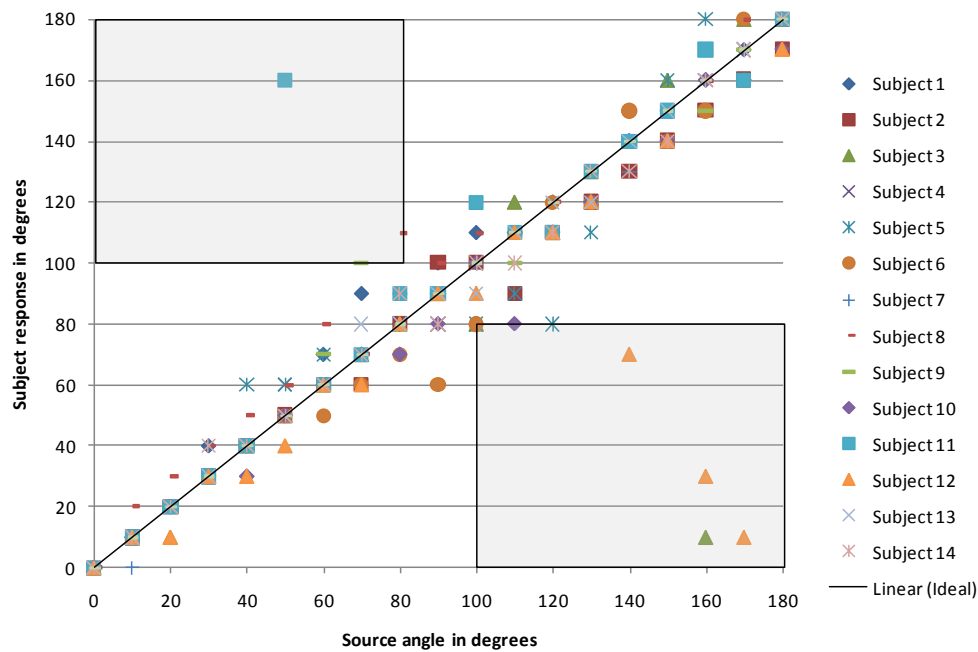


Figure 6-5: Subject response versus the actual source angle for low frequency noise

³ For clarity a front-back reversal is a user response that lies on the opposite side of the interaural axis from the actual sound source location.

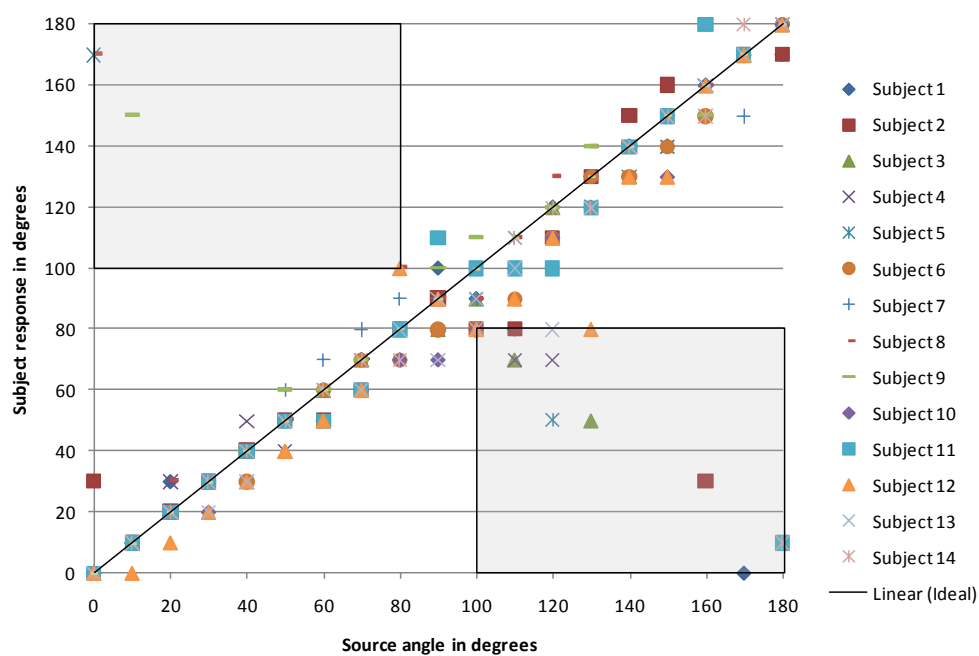


Figure 6-6: Subject response versus the actual source angle for mid/high frequency noise

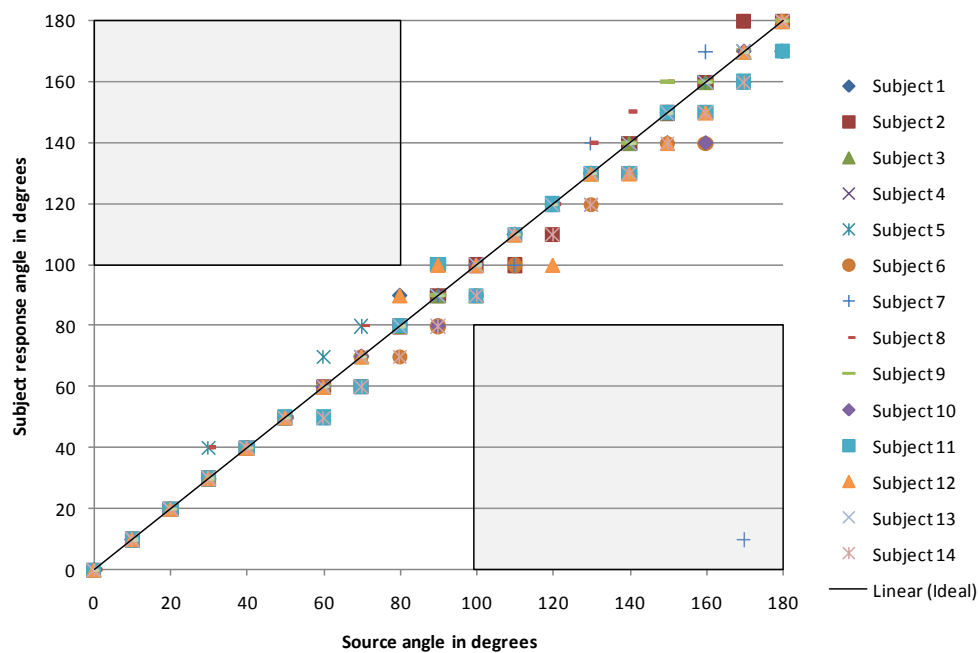


Figure 6-7: Subject response versus the actual source angle for male speech

In the low frequency noise test, approximately 5% of the total number of user responses was deemed to be a front-back reversal. For the mid/high frequency noise test, this increased to approximately 9%, whereas for male speech front-back reversals only accounted for less than 1% of all the user responses. Table 6.1 details the frequency of occurrence of front-back reversals by angle for each of the tests. Different cell shading is used for guidance.

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Low noise						1		1	1	-
Mid/high noise	2	1							2	-
Male speech										-
	100°	110°	120°	130°	140°	150°	160°	170°	180°	Σ
Low noise	4	1	1		1		2	1		13
Mid/high noise	5	4	3	2			1	1	2	23
Male speech								1		1

Table 6.1: Total number of front-back localisation confusions by source angle for each test.

Three interesting points can be drawn from the data in table 6.1. Firstly, front-back discrimination appears to be poorest in the region where localisation error is at its greatest (i.e. at the sides). This finding agrees with other research work (Simon R Oldfield & Parker 1984). The second interesting point is that there are more front-back confusions for mid/high frequency noise and low frequency noise than for male speech. The third point is that there are significantly more back to front confusions than there are front to back confusions – a point also observed in previous experiments (Simon R Oldfield & Parker 1984)

Front-back reversals require special treatment when determining the localisation error for each subject. For example, if a sound presented at 10° is reported by a subject to occur at 160° a straight subtraction to obtain the error would give -150°. This is clearly not an informative representation of the error. In order to overcome this problem the localisation error was adjusted in the reversed responses according to a method also used by Oldfield and Parker (Simon R Oldfield & Parker 1984). The adjustment entailed subtracting the sound source angle from 180° to give the sound source reversal position. Then the difference was taken between the sound source reversal position and the user's response. Using the example given previously, the

reversal method would produce an error of 10° instead of -150°. As expected, when applying this correction to the test data a much closer fit to the ideal response was attained.

To gain more information statistical analysis of the data was undertaken. The aim was to find out whether there were any significant differences between the localisation capabilities of each test subject and whether the type of sound source played a significant factor on subject performance. Investigation revealed that the most suitable statistical test to determine this was the Analysis of Variance (ANOVA). ANOVA is a powerful statistical technique that is used to test the hypothesis that the means from two or more groups of data are equal (Boslaugh & Watters 2008). It is commonly used in many different application areas when examining empirical data. One of the main parameters returned from an ANOVA test is the probability value p . The probability value indicates whether differences between the means under test are significant or not. Usually a value of less than 0.05 indicates that at least one sample mean is significantly different from the other sample means, whereas a value of greater than 0.05 generally indicates that none of the means are significantly different.

The first stage of the analysis consisted of running a 2-way mixed design ANOVA with subject and sound source being the two experimental factors. In this test the sound source was the repeated-measures variable and subjects was the between-subjects variable. For the repeated-measures both the original data and the reversal corrected data was used. During the test the statistical software (SPSS) checked the data for sphericity using Mauchly's test of sphericity (a requirement for a successful application of repeated-measures ANOVA). Where sphericity was not found the results are shown with the Greenhouse-Geisser correction, which avoids invalid calculations of F ratios and significance values p . Please note that this procedure was used for all repeated-measures ANOVA tests used in this chapter.

The results from this test show that both experimental factors were significant (i.e. sound source was $F(2,13, 537.54) = 11.55, p < 0.01, \eta_p^2 = .044$ and subject was $F(13, 52) = 1.95, p = 0.025, \eta_p^2 = .091$). Interestingly, the interaction between the factors was not deemed significant

$F(27.73, 537.54) = 1.15, p = 0.269, \eta_p^2 = .056$ suggesting that individual subject performance was not influenced by the sound source. The full results from this test are shown in table 1 of Appendix C.

To further investigate the effect of the sound source an additional 1-way between-subjects ANOVA was run. The results indicate that there is a significant difference between sound source for both the original data (i.e. $F(1.67, 442.70) = 8.50, p = 0.01, \eta_p^2 = .031$) and the reversal corrected data (i.e. $F(1.89, 499.88) = 9.85, p < 0.01, \eta_p^2 = .036$). As expected, the male speech source was the easiest to localise - this was declared by the majority of subjects when informally questioned after the test. The difference between the low frequency noise and mid/high frequency noise sources was not deemed significant. Both gave similar performance apart from the higher number of front-back reversals for mid/high frequency noise (see table 2 and table 3 of Appendix C for full ANOVA output).

In order to investigate the performance of the subjects further, a 1-way between-subjects ANOVA test of the subject localisation errors for all three real source localisation tests was run (the full results are provided in table 4 of Appendix C for the original data and the data with the reversal correction applied). For the original data the ANOVA results for the mid/high frequency noise and male speech tests indicate that there were no significant differences between the test subjects (i.e. $F(13, 252) = .306, p = .991, \eta_p^2 = .132$ and $F(13, 252) = 1.432, p = .145$). However, for the low frequency noise test, a significant difference was found between the capabilities of subjects when localising this source (i.e. $F(13, 252) = 3.366, p < 0.01, \eta_p^2 = .132$).

A multi-comparison test was carried out to identify which subjects were deemed significantly different for the low frequency noise source. The test revealed that the mean error for subject 12 was significantly different than subjects 1, 2, 4, 6, 7, 10, 13 and 14 (see figure 6-8) (please note that the error bars in this figure represent 95% confidence intervals - the means are deemed significantly different if their intervals are disjointed and are not significantly different when they overlap). It was found when examining the data that this particular subject experienced a larger

number of front-back reversals for this type of stimuli explaining the significance. After applying the reversal correction the ANOVA determined that there was no significant difference between the subjects for this source (i.e. $F(13, 252) = 1.544$, $p = .102$, $\eta_p^2 = .074$).

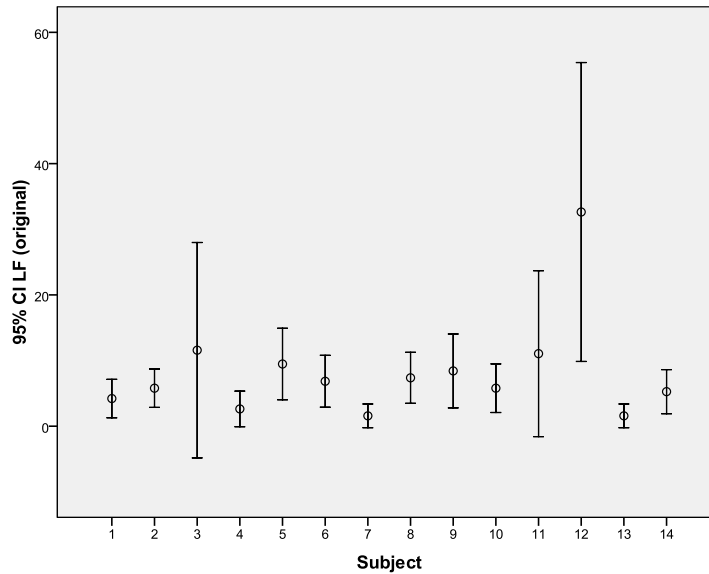


Figure 6-8: Multiple comparison test between the 14 subjects for the low frequency noise source. This is for the original listening test data with no front-back reversal correction applied.

The ANOVA results for the reversal corrected data show that there were no significant differences between subjects in the low noise and mid/frequency noise test (i.e. $F(13, 252) = 1.544$, $p = .102$, $\eta_p^2 = .074$ and $F(13, 252) = 1.528$, $p = .107$, $\eta_p^2 = .073$) but there was a significant difference between the subjects for the male speech stimulus, $F(13, 252) = 3.321$, $p < 0.01$, $\eta_p^2 = .146$.

A multiple comparison test revealed that the performance of subjects 6 and 14 were significantly different from subjects 3, 4, 9 and 13 (see figure 6-9). The perfect test scores for subjects 4 and 13, and the lower scores (in comparison) for subjects 6 and 14 is what defined the statistical significance flagged by the ANOVA test.

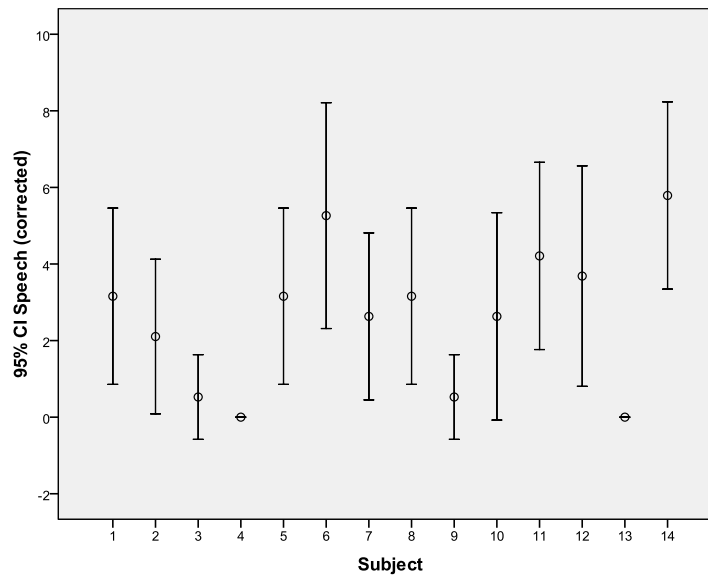


Figure 6-9: Multiple comparison test between the 14 subjects for the male speech source. This is for the test data with the reversal correction applied.

The reason why subject 6 and 14 did not perform as well for the male speech stimuli is unclear. In both the low frequency noise test and the mid/high frequency noise test (which are arguably more difficult sources to localise) both subjects performed relatively well on average. These subjects only took this experiment once so this point could not be investigated further. It is well known, however, that many individual factors can influence the accuracy of a subject's response during a listening test (e.g. tiredness and concentration).

Having considered the performance of all subjects in both sets of results (i.e. with and without reversals) the decision was made to include the data from all subjects when conducting further analysis. This decision was based upon the insight gained in the multiple comparison tests. Specifically, the error bars for the statistically worse performing subjects overlapped with the error bars for other subjects suggesting the differences in the capability of each subjects were not that great.

Figure 6-10, figure 6-11 and figure 6-12 plot the mean localisation error by angle with 95% confidence intervals for the respective sound sources. Each figure shows the original data (red line with square markers) and the data with the front-back reversal correction applied (blue line with circle markers). Note the higher mean error values at the front and rear for the mid/high frequency noise stimuli because of the front-back confusions.

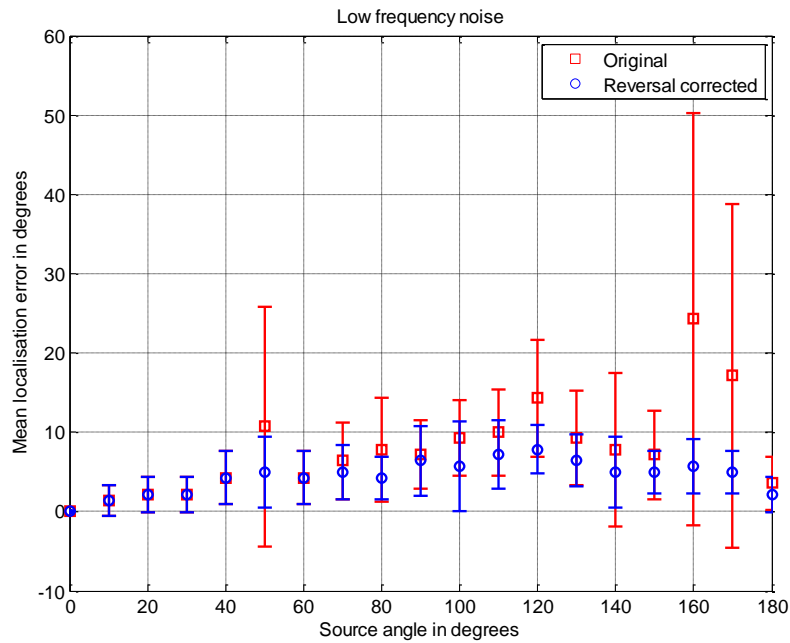


Figure 6-10: Mean localisation error by angle across all listening test subjects for the low frequency noise source with 95% confidence intervals.

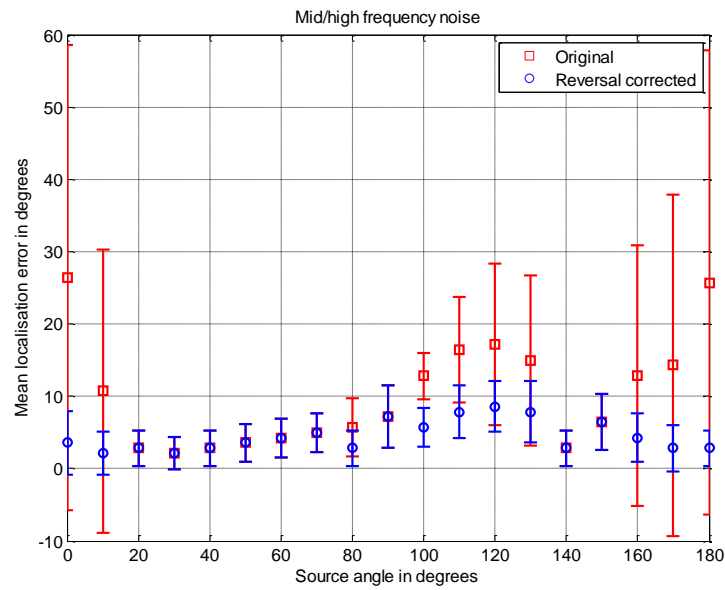


Figure 6-11: Mean localisation error by angle across all listening test subjects for the mid-high frequency noise source with 95% confidence intervals.

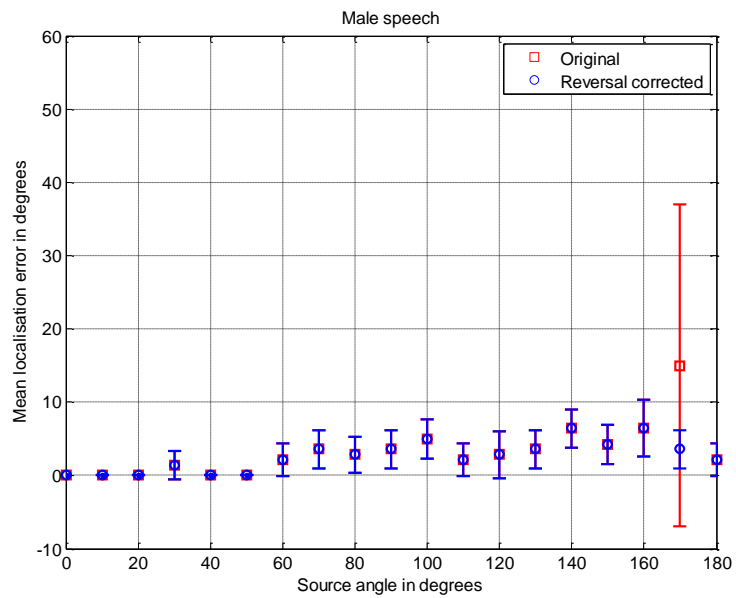


Figure 6-12: Mean localisation error by angle across all listening test subjects for the male speech source with 95% confidence intervals.

When considering the data with the front-back reversal correction applied, it is clear that localisation accuracy is best in the front and the direct rear but worse at the rear side. This overall trend correlates with existing psychoacoustic experimental data (Simon R Oldfield & Parker 1984; Blauert 2001). Without the front-back correction applied there are large mean localisation errors and 95% confidence intervals particularly at the front and rear for the noise sources. The mean and 95% confidence intervals show that the sound source angle was significant for all sources between the front and the rear-side.

6.4.3 Summary

Subjects were required to localise real sources of sound in the horizontal plane. The results from this test correlate with other research work and present a “best case” when evaluating the localisation performance of the developed decoders under the same listening conditions.

6.5 Test 2 - Decoded sound source localisation from the central listening position

6.5.1 Decoders under assessment

The following table provides information about each of the eight decoders used in this test.

Decoder	Order	Dual band	Comments
1	First	<input type="checkbox"/>	Derived using the original improved fitness function. This decoder was chosen because of its ideal velocity vector response (see figure 5.1).
2	First	<input checked="" type="checkbox"/>	Derived using the original improved fitness function. The only frequency dependent decoder employed in the test (see appendix B for implementation details).
3	First	<input type="checkbox"/>	Equivalent to the default settings on the Soundfield SP451 decoding unit. It was used as an anchor in this test because of the predicted poor performance (see figure). It has also performed badly in previous listening tests (Wiggins 2004).
4	Fourth	<input type="checkbox"/>	Derived using the improved fitness function with range-removal (see figure 5.5). This decoder was chosen because of its good overall predicted performance.
5	Fourth	<input type="checkbox"/>	Derived with the even error optimisation criteria included in the fitness function with range-removal. This decoder was chosen to represent an even performance decoder (see figure 5.8).
6	Fourth	<input type="checkbox"/>	Derived using the improved fitness function with range-removal. The MAA optimisation component switched on in the design tool (see figure 5.9).
7	Fourth	<input type="checkbox"/>	This decoder was derived by Wiggins (Wiggins 2007). The design tool performance plot of this decoder is included in appendix B.
8	Fourth	<input type="checkbox"/>	This decoder was derived by Craven (Craven 2003). The design tool performance plot of this decoder is included in appendix B.

Table 6.2: Information about the decoders used in the central listening point test.

As previously stated, the aim was to prove the concept of each component of the design tool. So, ideally, the outcome of this test would show the performance of the two first order decoder developed in this work (Decoders 1 and 2) to be better than Decoder 3 (default Soundfield).

Decoder 4 (standard fourth order) would ideally have performance which equals or exceeds the performance of Decoder 7 (fourth order by Wiggins) and Decoder 8 (fourth order by Craven). Decoder 5 (even error optimised) would ideally have even performance which is perceptible by the listeners. Decoder 6 (MAA optimised) would ideally have the best performance at the front of the ITU system out of all decoders at the cost of performance to the sides.

6.5.2 Test procedure

In this test subjects were assessed on their ability to localise decoded sound sources panned at intervals matching the real source localisation test (i.e. every 10°). Because of the sheer amount of potential evaluations (i.e. 19 different angles around the sound stage, 8 different decoders, 3 different stimuli), the number of angles assessed was reduced from 19 random angles to 10 random angles for each decoder. Each decoder was tested at each angle at least 3 times. In total each subject made 240 evaluations per test which took approximately 45 minutes to complete. Because of the time involved, subjects were advised to take a short break after completing 80 evaluations and 160 evaluations to reduce the risk of tiredness. The test was “double-blind” in that the presentation of the decoders and sound sources were randomised to the test subjects and test instructor.

6.5.3 Results

6.5.3.1 Front-back reversals

As expected, front-back reversals were apparent when localising all three sources of sound. Figure 6-13 shows the percentage of total user responses that were deemed front-back confusions for each decoder and for each sound source. The real source test data is included for comparison.

When taking into account the data from all decoders the sound source with the greater proportion of front-back reversals was the mid/high frequency noise source. This was followed by the low frequency noise then the male speech (agreeing with the data from the real source test).

Examining the decoders on an individual basis reveals that not all followed the same front/back reversal pattern mentioned above. For example, 4 out of the 8 decoders had a greater number of

front-back confusions for low noise than for mid/high noise (i.e. Decoders 1, 5 6 and 7). It was not obvious why this was the case.

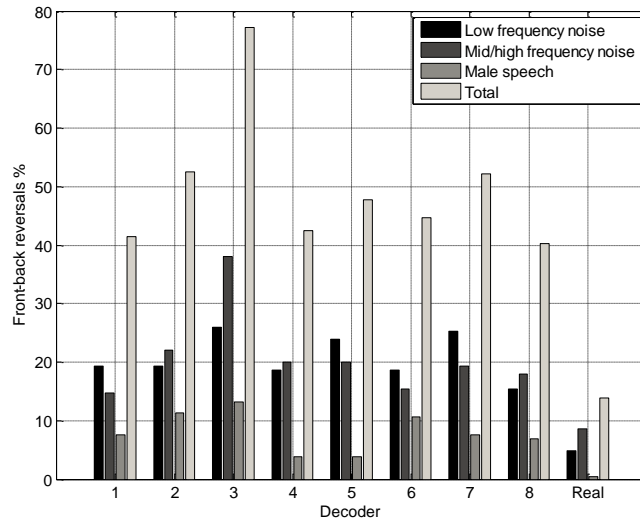


Figure 6-13: Percentage of front-back reversals for each decoder and for each sound source.

In total, Decoder 8 (fourth order decoder by Craven) gave the least number of front-back confusions. This was closely followed by Decoder 1 (first order frequency independent) and Decoder 4 (standard fourth order decoder). All other decoders gave comparable performance apart from the anchor (Decoder 3) which had considerably more front-back reversals.

Although front-back confusion is considered a natural limitation of the auditory system rather than the sound reproduction system, overall the percentages of front-back confusions were higher in this test than in the real source localisation test. Furthermore, subjects performed better for some decoders than for others. Because of this the following analysis will present the results with front-back reversals in the data and without.

6.5.3.2 Preliminary analysis

Preliminary analysis of the results was undertaken to identify the significance of the two main experimental factors (decoder and sound source). A 2-way repeated-measures ANOVA test was

run using both the original data and the reversal corrected data. The full results for these tests are reported in table 5 and table 6 of Appendix C.

For the original data (i.e. without the reversal correction) the ANOVA reported the difference between the performance of the decoders to be highly significant, $F(6.12, 972.54) = 24.21$, $p < 0.01$, $\eta_p^2 = .132$. This was also the case for the reversal corrected data $F(5.09, 809.76) = 51.44$, $p < 0.01$, $\eta_p^2 = .244$. This result will be investigated in greater detail in the sections following.

The ANOVA also confirmed a significant difference between sound sources (i.e. $F(1.71, 271.58) = 34.38$, $p < 0.01$, $\eta_p^2 = .178$ for the original data, and $F(2, 318) = 18.55$, $p < 0.01$, $\eta_p^2 = .104$ for the corrected data). Multi-comparison revealed the male speech was deemed the easiest source to localise in both cases, which was expected given the results found in test 1.

Interestingly, the ANOVA also revealed a significant interaction between decoder and sound source. This result indicates the level of performance for a decoder can change significantly according to the sound source used (i.e. $F(10.616, 1687.99) = 2.27$, $p < 0.01$, $\eta_p^2 = .014$ for the original data and $F(11.991, 1906.56) = 2.13$, $p = 0.013$, $\eta_p^2 = .01$) for the reversal corrected data. In order to illustrate this figure 6-14 and figure 6-15 presents the mean and 95% confidence intervals for each decoder and for each sound source for originals and reversal corrected data respectively (note the change in y axis scale between the two figures).

In both figures it is clear that there are cases where some decoders perform significantly better for one sound source than the other. For example, in the original data plot Decoder 1 and Decoder 2 perform significantly better for male speech than for mid/high frequency noise. This is also the case for the reversal corrected data.

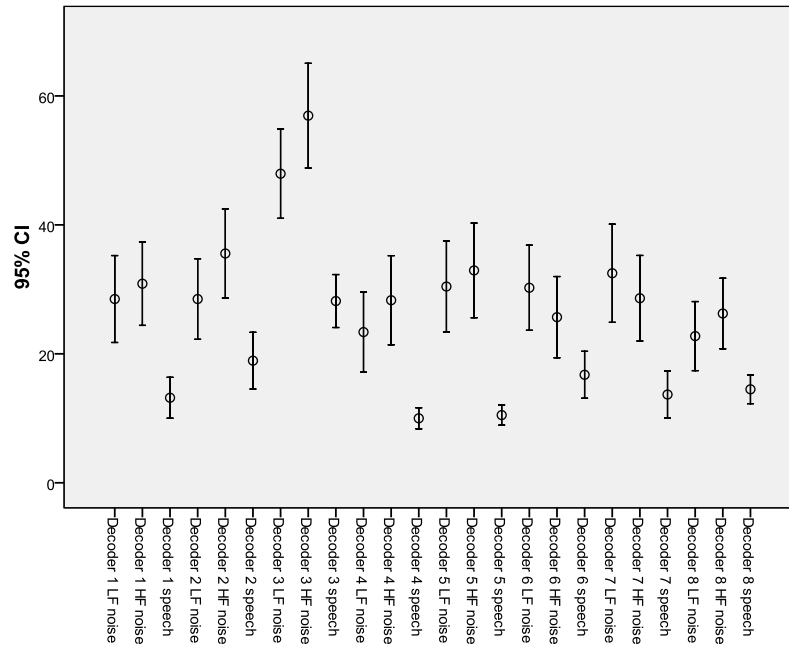


Figure 6-14: Mean and 95% confidence intervals for all decoders and all sound sources (original).

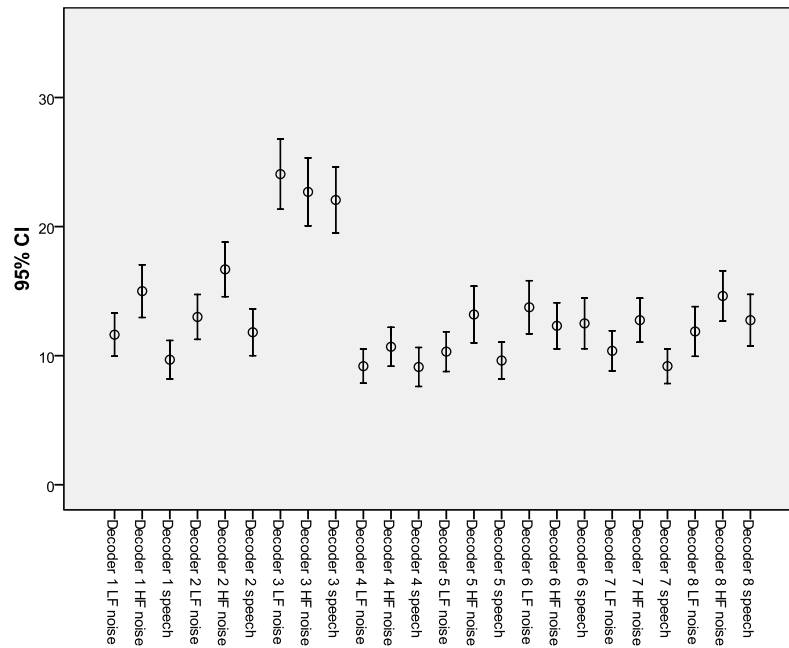


Figure 6-15: Mean and 95% confidence intervals for all decoders and all sound sources (reversal corrected).

6.5.3.3 Low frequency noise

Figure 6-16 presents the mean localisation error by angle for each decoder for the low frequency noise source. The data presented in this figure is the original data collected from the test (i.e. without the front-back reversal correction). For comparison, the equivalent data from the real source localisation test is included in each subplot.

The mean localisation errors of Decoder 1 (first order frequency independent) and Decoder 2 (first order frequency dependent) are closely matched. The only notable difference is that Decoder 2 has a slightly lower error at the rear of the system. Similar performance was anticipated for these decoders before the test because of their identical velocity vector responses.

Decoder 4 (standard fourth order) performed well with a strong correlation with the real source data between 0 degrees and 110 degrees. After 110 degrees the number of front-back reversals increases (see 170 degrees for example). Decoder 8 (fourth order decoder by Craven) also performed well with a strong correlation with the real source data between 40 degrees and 100 degrees. This decoder maintained a relatively low mean error at the rear because of the small number of front-back reversals in this area.

The mean localisation error for Decoder 5 (fourth order even error optimised) and Decoder 7 (fourth order by Wiggins) is comparable with the Decoder 4 and Decoder 8 at the sides of the listener. However, at the front both decoders have a higher mean error because of the greater number of the front-back reversals.

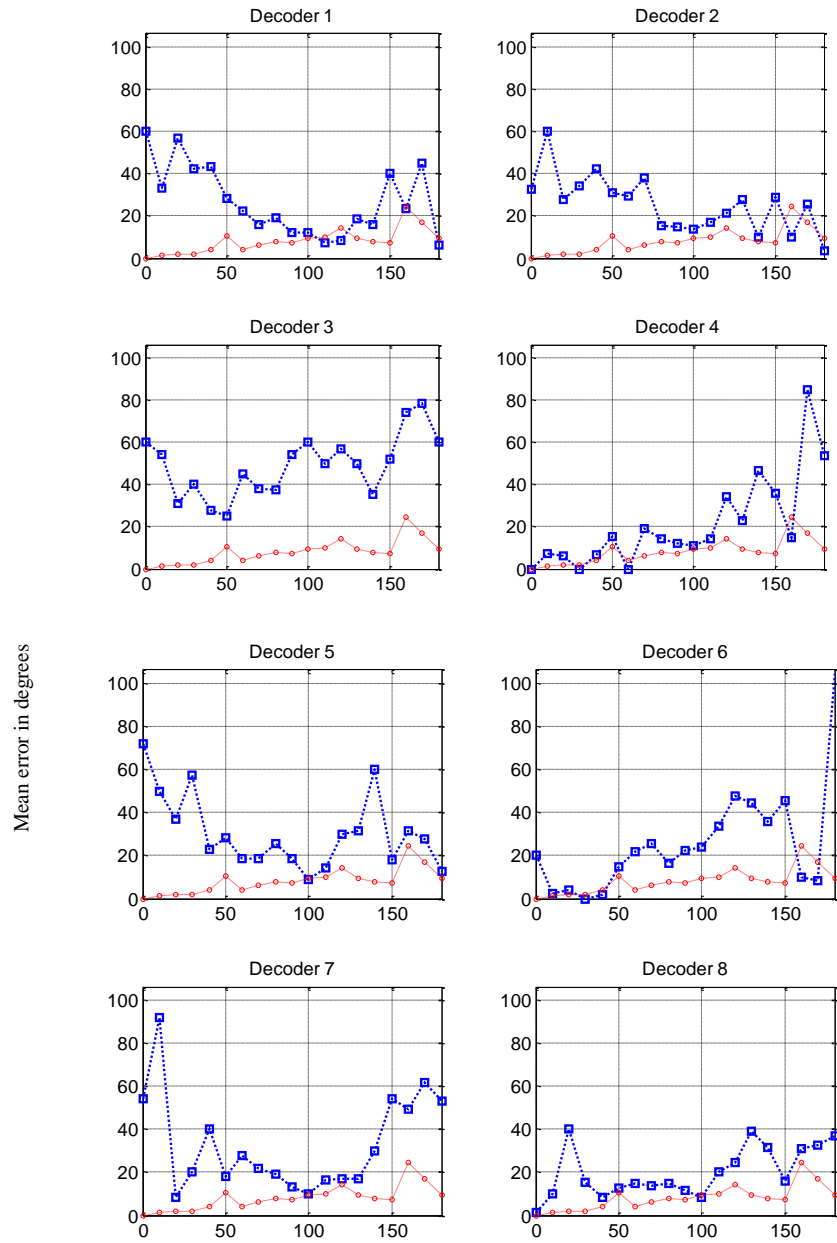


Figure 6-16: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the low frequency noise test (blue line). The equivalent data from the real source localisation test is included for comparison (red line). This is the original data without the front-back reversal correction applied.

As desired, Decoder 6 (MAA optimised) performed well at the front of the system between 10 degrees and 50 degrees. The data for this decoder had the closest match and strongest correlation with the data from the real source test in this area. The anchor (Decoder 3) gave the highest mean error by angle overall. The level of error for this decoder is consistently high around the sound stage.

A 1-way repeated-measures ANOVA was used to investigate whether there was any significant difference between the performances of the eight decoders for this particular scenario. The full ANOVA results are presented in table 7 of Appendix C.

In this case the ANOVA results show there was a significant difference between the performance of the decoders (i.e. $F(5.74, 913.26) = 8.358, p < 0.01, \eta_p^2 = .050$). To illustrate this, figure 6-17 shows the overall mean error with 95% confidence intervals for each decoder. This data suggests that Decoder 4 (standard fourth order) and Decoder 8 (fourth order by Craven) are potentially the best in this scenario although this is not statistically proven because of the overlapping confidence intervals. The only thing that is proven is that all decoders, apart from Decoder 7 which suffered from a high number of front-back reversals, were significantly better than Decoder 3.

Please note that the standard deviation of the mean localisation error by angle was included in table 6.3 as a measure of even performance. However, when examining the original data the standard deviation value for each decoder is strongly influenced by the number of front-back reversals.

Decoder:	1	2	3	4	5	6	7	8
Standard deviation	16.75	13.37	14.71	21.74	17.42	24.59	22.28	11.70
Mean difference from real source	18.65	17.26	40.71	12.85	22.49	17.37	24.64	11.99

Table 6.3: Standard deviation of the mean localisation error by angle and the mean difference from the equivalent real source data for the low frequency noise test (original data).

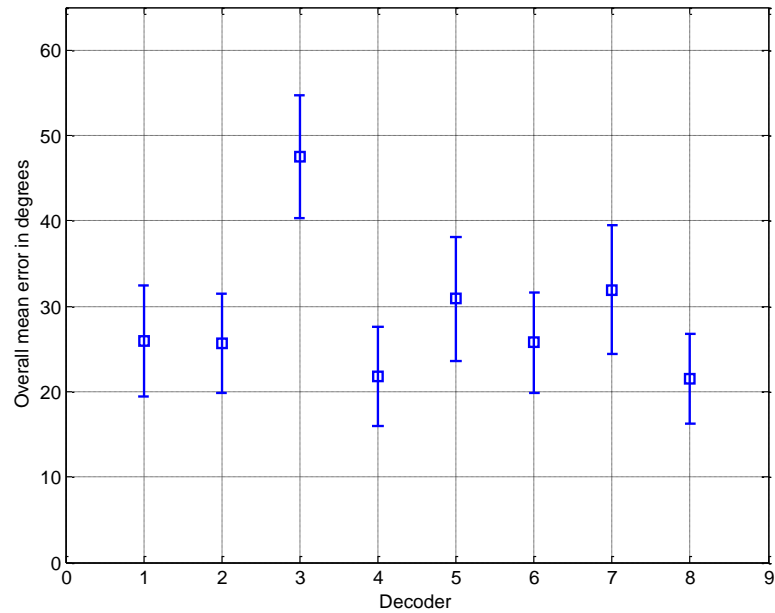


Figure 6-17: Overall mean error with 95% confidence intervals for each decoder from the low frequency noise test (original data).

Figure 6-18 presents the mean localisation error by angle for each decoder for the low frequency noise source with the front-back reversal correction applied. As expected, removing the reversals has resulted in lower mean localisation errors at the front and the rear of the system for some decoders (please note a smaller y-axis scale was used in this figure when compared to figure 6-16 because of the lower mean errors).

Interestingly, two of the first order decoders (Decoder 1 and 2) had ideal velocity vector responses but neither outperformed the fourth order decoders which did not have ideal velocity vector responses at the sides and rear of the system (this was also true of the original data for this sound source test). This implies that other perceptual factors are important when producing 5-speaker decoders with good low frequency performance. One factor that may influence performance is that first order decoders play sound out of all loudspeakers simultaneously at a greater level than the higher order Ambisonic decoders, possibly leading to biases in the listener's response.

When comparing all of the decoders in figure 6-18 it is clear that Decoder 4 (standard fourth order) is the closest match to the real source data. This decoder performed consistently well, even through the area of the system where human localisation ability is at its poorest (rear-side).

As desired, the MAA optimised decoder had the best frontal performance. Between 0° and 40° listeners found phantom sound sources produced by this decoder the easiest to localise. Note the performance to the sides for this decoder is also reduced when compared to the front and rear suggesting that the MAA optimisation component introduced into the design tool was successful in this frequency region.

Decoder 3 produced the highest mean localisation error by far at the side of the listener (note that the error builds to a peak at the angle of 90 degrees). This suggests that listeners found phantom sound sources very difficult to perceive in this area of the system.

A repeated-measure ANOVA revealed that there were significant differences in terms of localisation performance between the decoders when examining the low frequency noise test data with the reversal correction, $F(5.44, 865.21) = 27.55, p < 0.01, \eta_p^2 = .148$ (the full ANOVA results are provided in table 8 of Appendix C). Overall Decoder 4 (standard fourth order) performed significantly better than Decoders 2, 3, 6 and 8 (illustrated in figure 6-19).

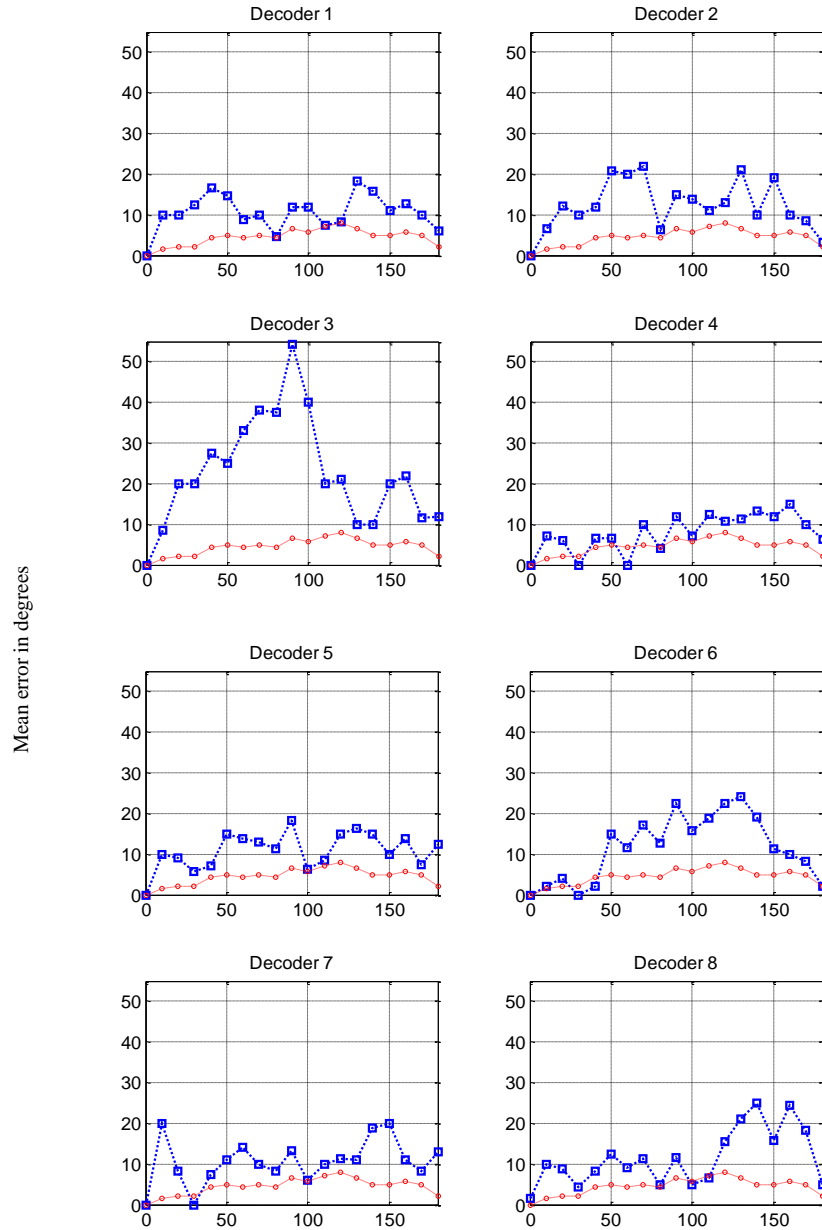


Figure 6-18: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the low frequency noise test (blue line). The equivalent data from the real source localisation test is included for comparison (red line). The data presented in this figure includes the front-back reversal correction.

Decoder:	1	2	3	4	5	6	7	8
Standard deviation	4.35	6.19	13.30	4.57	4.48	8.08	5.50	6.86
Mean difference from real source	6.40	7.88	18.18	3.48	6.47	7.08	6.15	7.06

Table 6.4: Standard deviation of the mean localisation error by angle and the mean difference from the equivalent real source data for the low frequency noise test (with reversal correction).

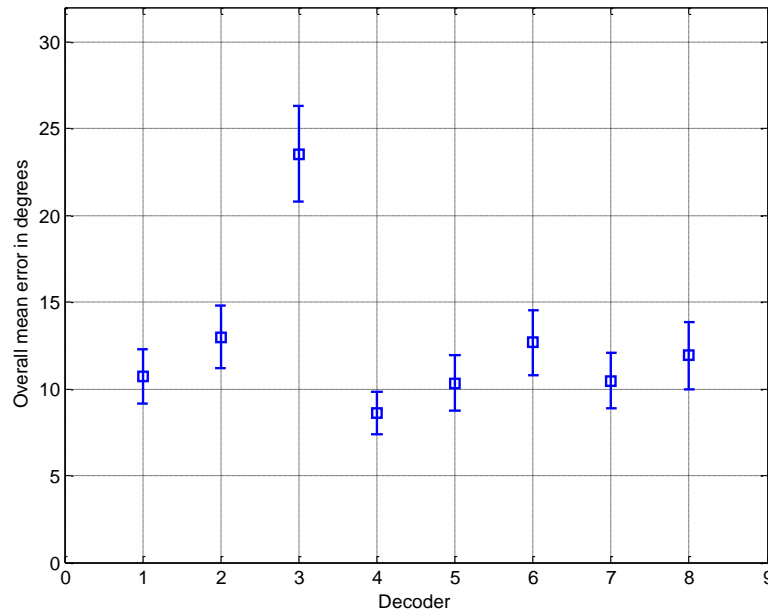


Figure 6-19: Overall mean error with 95% confidence intervals for each decoder from the low frequency noise test (with the front-back reversal correction).

The standard deviation values presented in table 6.4 as a measure of even performance show that the localisation error for Decoder 1 (first order frequency independent) was the most consistent around the sound stage. Relatively low standard deviation values for Decoder 4 (standard fourth order) and Decoder 5 (even performance optimised decoder) were also achieved for this source.

Both sets of results (with and without the reversal correction) confirm all decoders were significantly better than the anchor (Decoder 3) when examining the low frequency noise source data.

6.5.3.4 Mid/high frequency noise

Figure 6-20 presents the mean localisation error by angle for each decoder for the mid/high frequency noise source test. The data presented in this figure is the original data collected from the subjects (i.e. without the reversal correction).

The error trend of Decoder 4 (standard fourth order) has the strongest correlation with the data from the real source localisation test. However, this decoder has a highest mean localisation error at 0 degrees and 180 degrees because of front-back reversals.

Decoders 6, 7 and 8 have comparable mean localisation error responses though the mean localisation error for Decoder 6 is slightly lower than the other decoders around the front of the system (between 30 and 60 degrees) and more consistent around the front-side. Improved performance at the front was a desired performance characteristic for Decoder 6 (MAA optimised decoder).

Decoder 5 (even error optimised) appears to produce relatively good performance at the direct side. However, performance appears is reduced at the front and the rear. As mentioned previously, even performance by angle is difficult to assess with front-back reversals in the data. Note the high standard deviation of the error for all decoders in table 6.5.

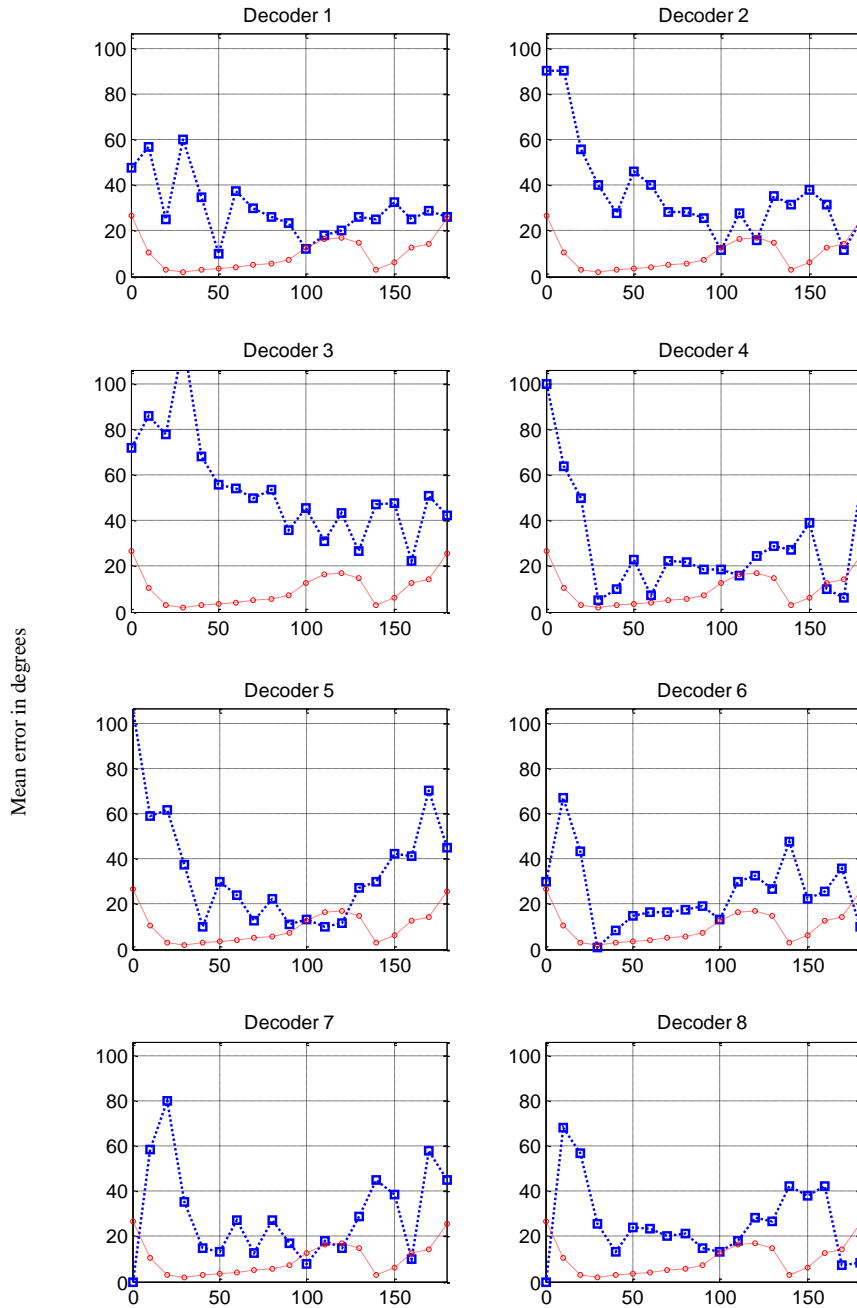


Figure 6-20: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the mid/high frequency noise test (blue line). The equivalent data from the real source localisation test is included for comparison (red line). This is the original data without the front-back reversal correction.

Decoder:	1	2	3	4	5	6	7	8
Standard deviation	13.17	21.69	22.96	24.25	25.57	15.66	20.75	17.13
Difference from real source	19.50	26.59	43.95	18.84	24.82	14.89	18.80	15.63

Table 6.5: Standard deviation of the mean localisation error by angle and the mean difference from the equivalent real source data for the mid/high frequency noise test (original data).

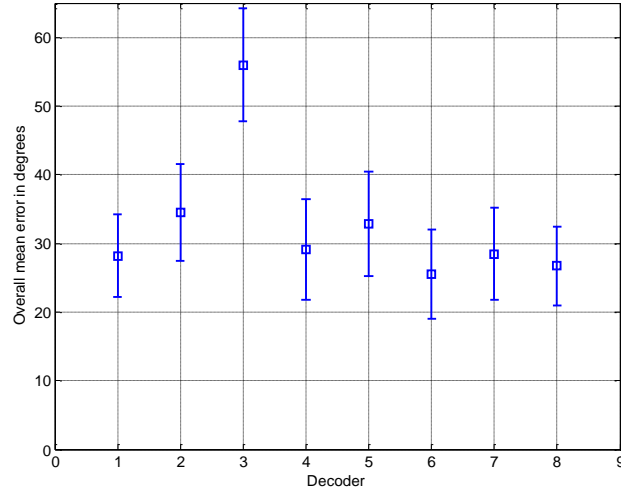


Figure 6-21: Overall mean error with 95% confidence intervals for each decoder from the mid/high frequency noise test (original data).

A 1-way repeated-measure ANOVA test flagged a significant difference between the performance of the decoders (i.e. $F(6.18, 983.27) = 11.24$, $p < 0.01$, $\eta_p^2 = .066$) (table 9 of Appendix C displays the full ANOVA data). Overall, Decoder 6 (fourth order MAA optimised decoder) has the lowest overall mean error in figure 6-21. However, the overlapping confidence intervals prevent this decoder from being significantly better when analysing the data in this way. What is shown in this figure though is that all decoders are significantly better than Decoder 3.

Figure 6-22 presents the mean localisation error by angle for the mid/high noise stimuli with the front-back reversal correction applied. When comparing this figure with figure 6-20 the impact of the front-back reversals in the data becomes clear. Front-back reversals are particularly visible in each subplot. Generally, for this sound source there were a greater number of front-to-back confusions than there were back-to-front confusions (note the high mean localisation errors at the front of the system).

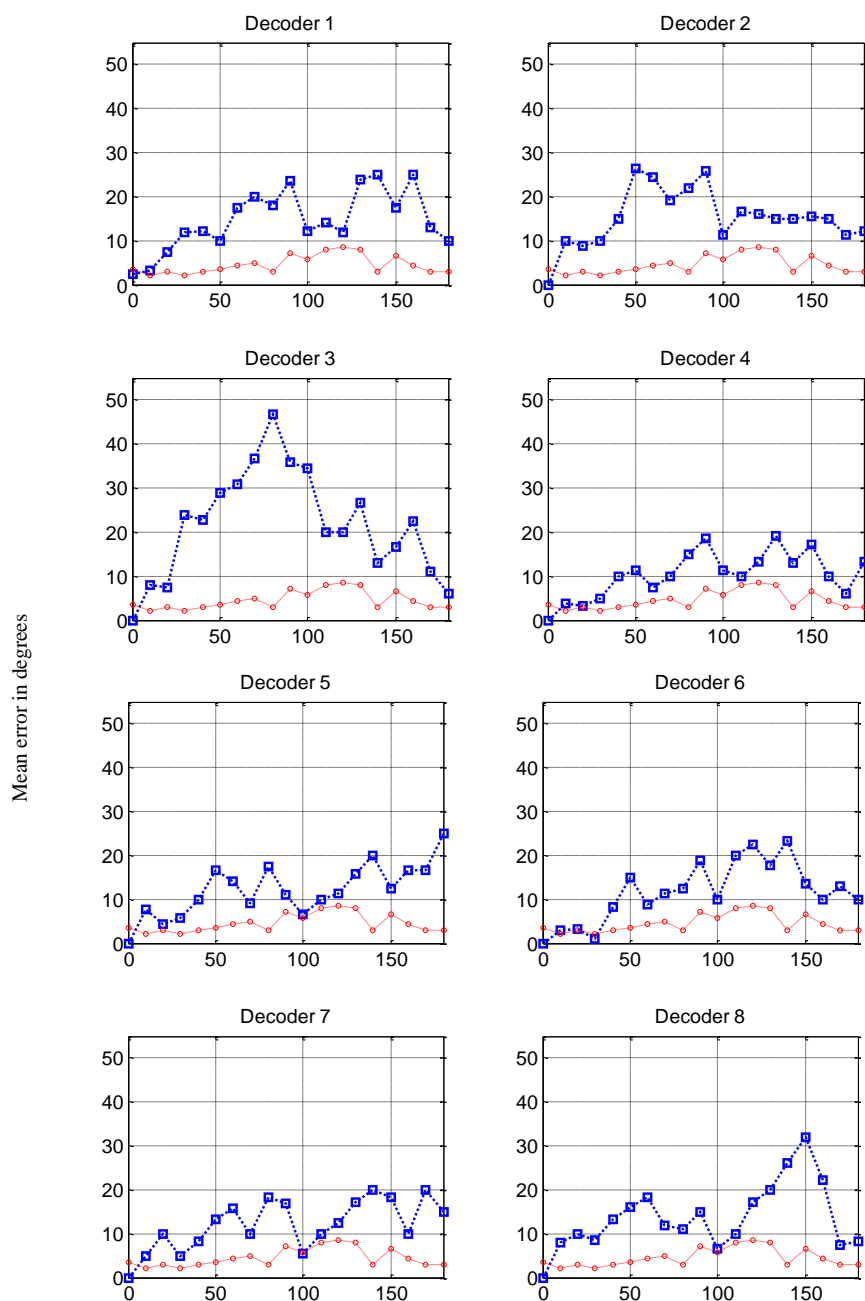


Figure 6-22: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the mid/high frequency noise test (blue line). The equivalent data from the real source localisation test is included for comparison (red line). The data presented in this figure has the front-back reversal correction applied.

Decoder:	1	2	3	4	5	6	7	8
Standard deviation	6.80	6.43	12.27	5.24	6.02	6.88	5.75	7.60
Difference from real source	10.18	10.74	17.14	5.89	7.63	7.18	7.65	9.29

Table 6.6: Standard deviation of the mean localisation error by angle and the difference from the equivalent real source data for the low frequency noise test (with the reversal correction).

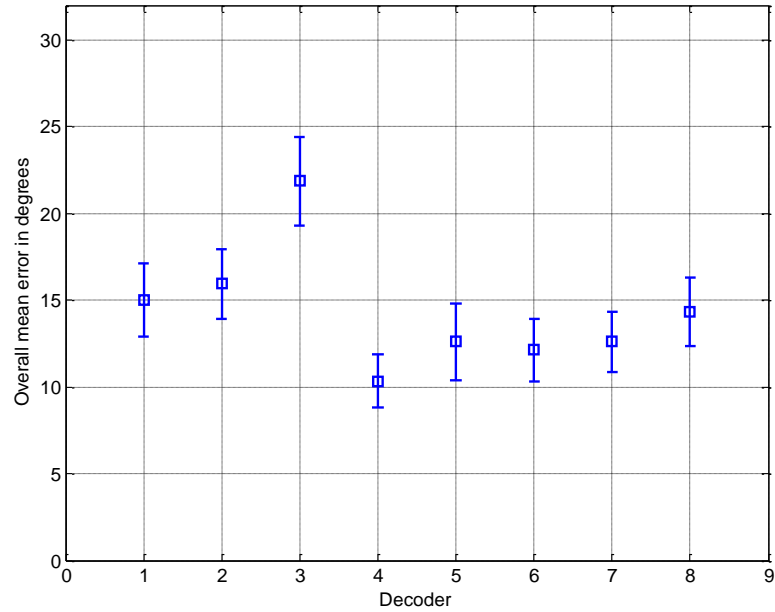


Figure 6-23: Overall mean error with 95% confidence intervals for each decoder from the mid/high frequency noise test (with the front-back reversal correction).

The first order decoders (Decoder 1, 2 and 3) have a higher mean error localisation error by angle than the fourth order decoders (as predicted by the energy vector). These decoders also have a higher overall mean error (see figure 6-23). Decoder 4 (standard fourth order) and Decoder 6 (MAA optimised) have the strongest correlation with the real source data. These decoders were ranked first and second respectively when considering their overall mean error values).

A repeated-measure 1-way ANOVA test was undertaken to determine the significance of the results in this particular test case (see table 10 for the full ANOVA output). The results show that

there was a significant difference between the performance of the decoders (*i.e.* $F(6.12, 973.51) = 11.24, p < 0.01, \eta_p^2 = .088$). Figure 6-23 shows the mean overall localisation error for Decoder 4 is significantly lower than Decoders 1, 2, 3 and 8. The mean overall localisation error for Decoder 6 is significantly lower than Decoder 2 and Decoder 3. All other decoders are significantly better than the anchor (Decoder 3).

This time the standard deviation values (see table 6.6) show that Decoder 4 had the most consistent mean error by angle. Decoder 5 (even error optimised) gave the most consistent performance between 0 degrees and 90 degrees as predicted.

6.5.3.5 Male speech

Figure 6-24 presents the mean localisation error by angle for each decoder for the male speech source. All data presented in this figure is the original data collected from the subjects in the test. The equivalent data from the real source test is included for comparison. Note the lower mean errors in general for this source especially at the front and rear because of the lower number of front-back reversals than the previous two tests.

This figure shows quite clearly that the localisation performance of Decoder 4 (standard fourth order decoder) was consistently good around the listener. The low standard deviation value presented in table 6.7 highlights this. This decoder was the closest match to the real source data and was able to maintain good performance when positioning sound sources to the sides and to the rear of the listener.

Decoder 5 (even error optimised decoder) also gave consistently good localisation performance (even through the problematic localisation area between 120° and 170° where it actually performed better than Decoder 4 on average). This is an interesting result given that this decoder did not perform as well as the other fourth order decoders in terms of velocity vector magnitude or energy vector magnitude. This decoder did, however, perform well for the vector angle objectives and out of all decoders gave the closest match for the angle match objective ($E_{AngMatch}$).

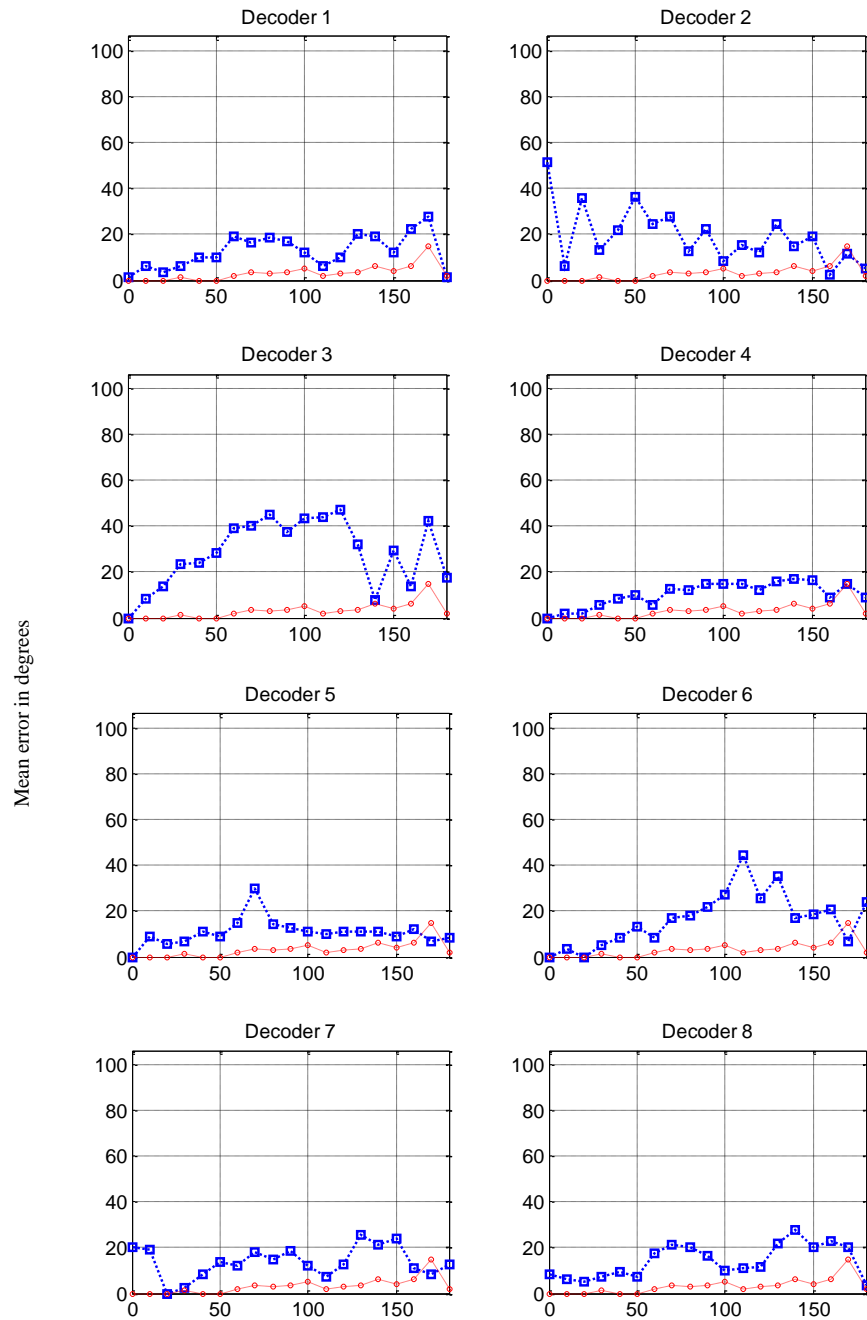


Figure 6-24: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the male speech test (blue line). The equivalent data from the real source localisation test is included for comparison (red line). This is the original data without the front-back reversal correction.

Decoder:	1	2	3	4	5	6	7	8
Standard deviation	7.49	12.37	14.60	5.36	5.77	11.77	6.84	7.16
Difference from real source	9.37	15.93	24.95	7.19	7.47	13.30	10.49	10.79

Table 6.7: Standard deviation of the mean error by angle and the mean difference from the equivalent real source data for the male speech test (original data).

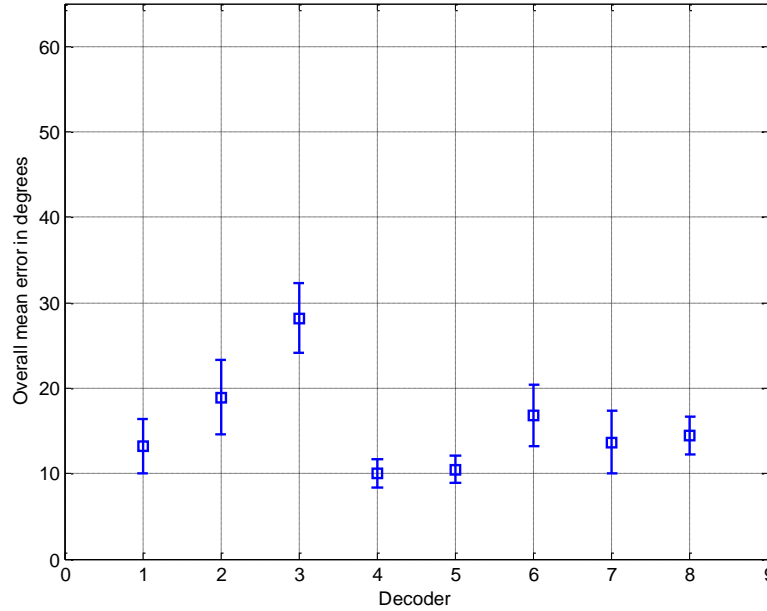


Figure 6-25: Overall mean error with 95% confidence intervals for each decoder from the male speech test (original data).

The even error optimised decoder also gives even performance around the listener as desired. This is highlighted by the relatively low standard deviation values in table 6.7. If we ignore the blip at 70 degrees and draw an imaginary line through the data we see very even performance.

Decoder 6 (MAA optimised decoder) also had a trend that matched its predicted localisation performance (i.e. good at the front at the cost of performance to the sides).

A repeated-measure 1-way ANOVA test was undertaken to investigate the significance of the results for this test case (see table 11 of Appendix B for full ANOVA output). The results confirmed that there was a significant difference between the performance of the decoders (i.e. $F(5.10, 810.74) = 13.75, p < 0.01, \eta_p^2 = .080$). Figure 6-25 illustrates this by plotting the overall means with 95% confidence intervals. The figure shows that Decoder 4 and Decoder 5 were very closely matched giving the lowest overall mean errors and that both decoders were significantly better than Decoders 2, 3, and 6.

Figure 6-26 presents the mean localisation error by angle for each decoder with the front-back reversal correction applied (please note the change in scale from previous mean error figures). The equivalent data from the real source test is included for comparison.

Decoder 5 (even error optimised) gave the most consistent mean error around the listener (note the low standard deviation value in table 6.8). The blip at 70 degrees noted previously was due to a front-back reversal.

The remaining localisation error trends presented in figure 6-26 are quite similar to those presented in figure 6-24 but with lower mean errors at the front and rear because of the removal of the front-back confusions.

A repeated-measure 1-way ANOVA test was run to determine any significant differences between the performances of the decoders in this test case (see table 12 of Appendix C for full ANOVA output). The results show that there was a significant difference $F(5.68, 903.046) = 25.79, p < 0.01, \eta_p^2 = .140$. The overall mean errors and confidence intervals in figure 6-27 confirm that Decoder 4 (standard fourth order) and Decoder 7 (fourth order by Wiggins) were significantly better than the Decoder 3 and 8. Decoders 5 and 1 were closely matched almost equalling Decoders 4 and 7. All decoders were significantly better than the anchor (Decoder 3).

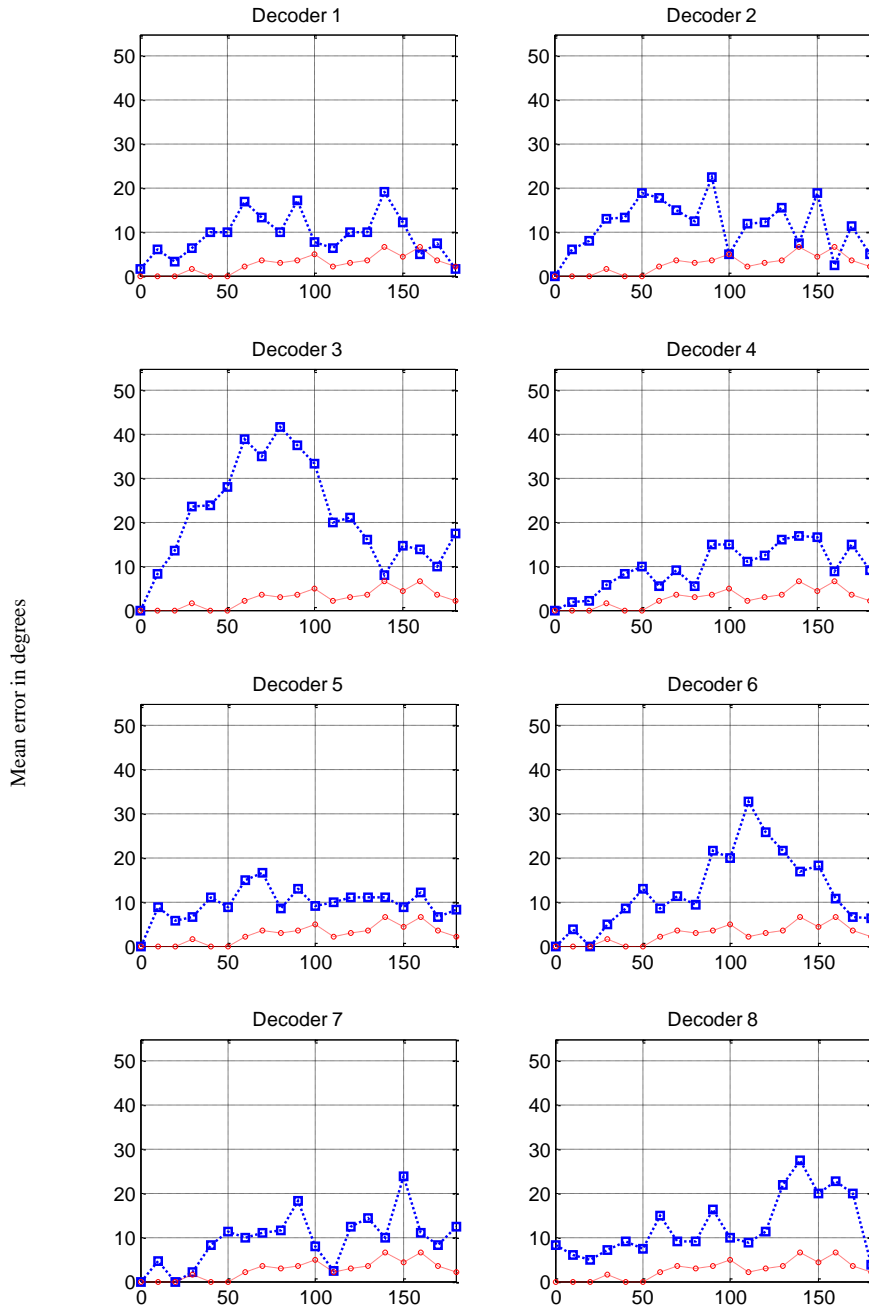


Figure 6-26: Mean localisation error by angle for each decoder taking into account the responses from all subjects in the male speech test (blue line). The equivalent data from the real source localisation test is included for comparison (red line). The data presented in this figure has the front-back reversal correction applied.

Decoder:	1	2	3	4	5	6	7	8
Standard deviation	5.04	6.06	11.81	5.29	3.60	8.92	6.00	6.90
Difference from real source	6.52	8.77	18.67	7.06	6.98	10.02	6.85	9.92

Table 6.8: Standard deviation of the mean error by angle and the mean difference from the equivalent real source data for the male speech test (with the reversal correction).

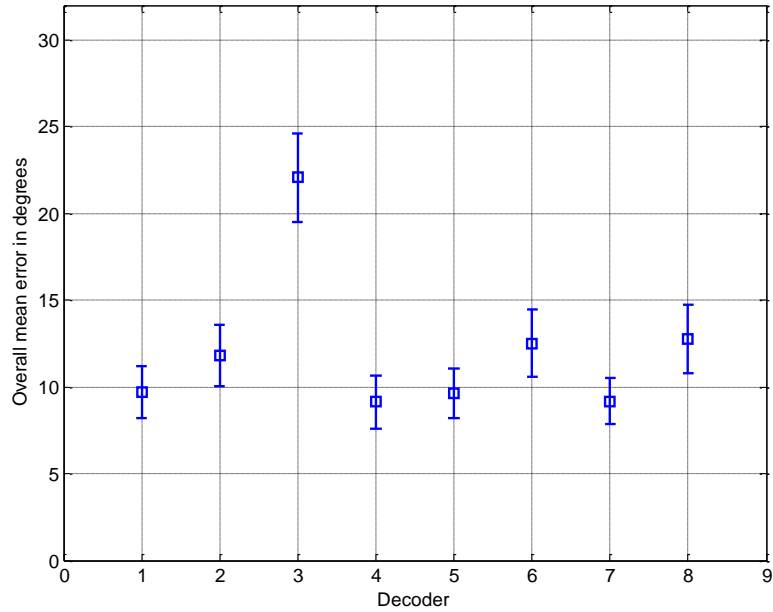


Figure 6-27: Overall mean error with 95% confidence intervals for each decoder from the male speech test (with the front-back reversal correction).

6.5.3.6 Overall decoder performance

In order get an idea of the best performing decoder across all tests an overall mean localisation error was derived⁴. Figure 6-28 shows the overall mean localisation error with 95% confidence intervals for each decoder taking into account the three tests. The plot shows the original data and the data with the front-back reversal correction.

⁴ We acknowledge that combining the data across tests when the sound source is deemed a significant factor is not a valid approach statistically.

Figure 6-28 suggests that Decoder 4 (standard fourth order decoder) performed the best. When examining the original data without the reversal correction this decoder gave a lower overall mean error than all other decoders. The only significant difference in terms of performance was for the anchor (Decoder 3) which was the worse overall as expected. When examining the data with the reversal correction applied Decoder 4 performed significantly better than Decoders 1, 2, 3, 6 and 8. In the majority of the test cases this decoder gave excellent frontal imaging and was able to give the closest match to the results obtained in the real source localisation test.

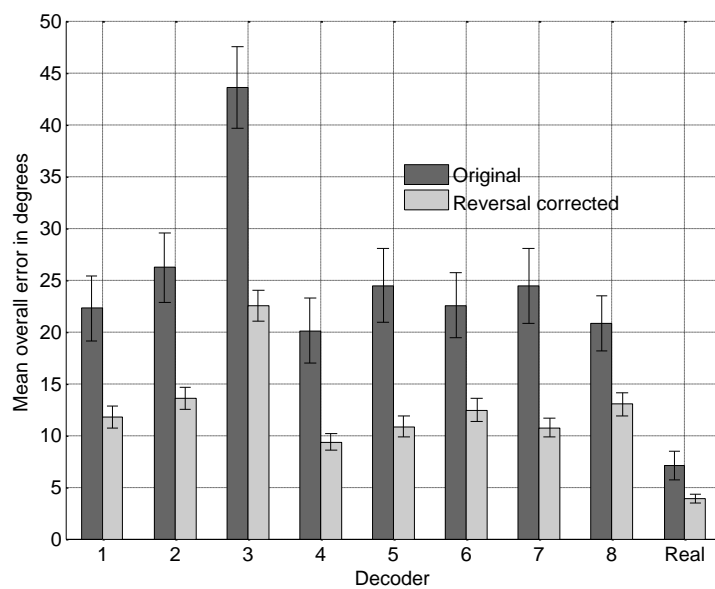


Figure 6-28: Overall mean localisation error with 95% confidence intervals for each decoder taking into account the data from the three sound source tests (original and reversal corrected).

6.5.4 Discussion

Taken as a whole, test subjects found the noise bursts the most difficult to localise and the male speech the easiest (correlating with the real source localisation test). This was shown by the mean error values and respective high and low standard deviation for these sources. When questioned after the test, individuals reported the noise bursts as sometimes being “elevated” above the horizontal plane when panned to the side. The mid/high frequency noise source was also reported as occasionally sounding “more distant” at the rear-side. It was not clear if these

particular events were as a result of source interaction with the room or whether it was due to the decoders.

Subjects commented that the male speech source was the easiest to localise because it was more “focused” and “longer in duration”. We acknowledge that small involuntary head movements may have helped listeners when localising this particular source as the restriction of head movements was not strictly enforced during the test.

When considering the results from all three sound sources it was the standard fourth order decoder (Decoder 4) that performed the best overall. This decoder performed relatively well in all cases presented either exceeding or approximately equaling the performance of the other decoders. The only exception to this was for mid/high frequency noise test (with reversals).

For all sound sources improved localisation was predicted and observed for Decoder 6 (MAA optimised) at the front of the 5-speaker system at the cost of performance to the sides. This meets the desired performance characteristics.

Mixed results were obtained with respect to the even error optimised decoder (Decoder 5). For the male speech stimuli very consistent localisation performance around the listener was observed (ignoring the noted anomaly at 70 degrees in the original data plot). However, for the noise bursts this was not always the case because of the high standard deviation of localisation error across the subjects. Further listening tests with easier to localise sound sources might reveal more information with respect to even performance.

Although the results presented for this test are encouraging, they are not completely conclusive. The first order decoders (excluding Decoder 3) both gave an ideal velocity vector response but this was not reflected in the results for the low frequency noise test. This prompts the question as

to whether other perceptual factors can influence a listener's judgment. Could the high level of loudspeaker crosstalk for the first order decoders be a problem for a centrally seated listener?

Importantly, the decoders selected for this test were produced using specific sets of importance weightings. They might not necessarily represent the best decoders the tool could produce for each specific scenario. As highlighted earlier, the advantage of using the design tool is the decoder designer could experiment with the tool adjusting importance weightings and auditioning different decoders. For example, if after auditioning one decoder a specific problem was identified, the designer could adjust the decoder tool's criteria and importance of different objectives to attempt to reduce these perceived weaknesses.

6.6 Test 3 - Decoded sound source localisation from off-centre listening positions

6.6.1 Decoders under assessment

The following table details the five frequency independent decoders evaluated in this test:

Decoder	Order	Comments
1	First	Decoder 1 from the previous listening test (the only first order decoder evaluated). It was expected to give the worst performance overall because of the large secondary lobes of the virtual microphones (see figure 5.1). Not optimised for off-centre performance.
2	Fourth	Decoder 4 from the previous test. This decoder was selected as it represents a good fourth order decoder optimised for the central listening point. Not optimised for off-centre performance.
3	Fourth	Derived by the design tool with the off-centre optimisation component switched on (see section 5.8).
4	Fourth	This decoder was derived by Wiggins and was included for comparison (appendix B shows the performance plot for this decoder). Not optimised for better off-centre performance.
5	Fourth	Derived by Poletti using different off-centre optimisation criteria (Poletti 2007) (see appendix B for more information).

Table 6.9: Information about the decoder used in the off-centre listening test.

The ideal outcome would be for the off-centre optimised decoder to reduce loudspeaker bias and to produce localisation performance equaling or exceeding the performance of the other fourth order off-centre optimised decoder designed by Poletti at all listening positions.

6.6.2 Test procedure

In this test subjects were required to localise panned sources from 9 different listening positions (figure 6-29 shows the arrangement of loudspeakers and the listening positions). The X and Y offset from each of the listening positions was 70cm giving listeners adequate room when seated.

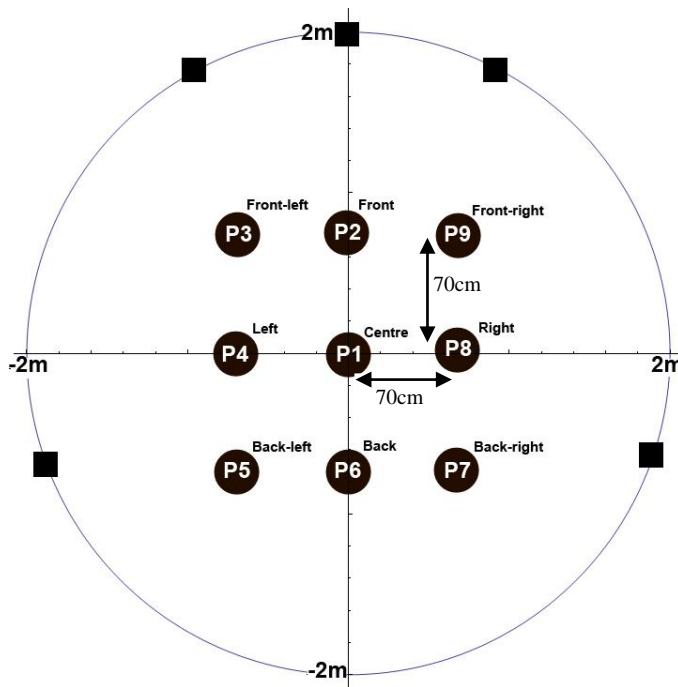


Figure 6-29: Listening positions (P1 to P9) which were evaluated in the off-centre test.

During the test sound was only presented to the listeners on the right side of the system. However, source angles were marked every 10 degrees on both sides of the system to allow for the possibility of sound being perceived as coming from the left side of the system when in positions P3, P4 and P5 (i.e. loudspeaker bias). Perceived angles on the left of the system were marked as negative angles.

The reader is reminded that the localisation error from the left side listening positions (P3, P4, P5) will be different to the right side listening positions (P7, P8, P9) when testing on one side of the listener. As will the localisation error from the front and rear listening positions (P2 and P6). In each position a listener has a different perspective of a reproduced sound source image.

When designing this test it was clear that asking subjects to evaluate the same number of source angles as in the central listener test would not be feasible. For example, evaluating 10 angles, for 5 decoders, for each listening position, and for each sound source would require a total of 1350

evaluations! In the light of this, the number of angles assessed was reduced to 7 random angles for each decoder (i.e. 35 evaluations for each listening position). After an initial trial, it was decided that subjects could manage to evaluate one sound source at 6 positions during one test sitting without risking tiredness (i.e. 210 evaluations).

The tests were conducted in groups of six and groups of three. In groups of six, the centre line and right side positions were simultaneously evaluated by each of the subjects (i.e. P1, P2, P7, P8 and P9). In the groups of three, the subjects simultaneously evaluated the left side positions (P3, P4 and P5). The former test took approximately 30 minutes to complete, whereas that latter test took approximately 15 minutes to complete. Subjects recorded their responses on paper. The test was “double-blind” in that the presentation of the decoders and sound sources were randomised to the test subjects and test instructor.

6.6.3 Results

The results will be presented without the front-back reversal correction (i.e. the raw experimental data). This is because in off-centre positions the distinction between a front-back reversal and a loudspeaker nearfield bias is difficult to determine. As a consequence, the errors for some decoders will be greater at the front and rear of the system when evaluated from the centre listening position (P1) than in the previous test.

6.6.3.1 Preliminary analysis

The results were analysed using a multi-way repeated-measures ANOVA with the main variables being decoder, sound source and listening position (table 13 of Appendix C provides the full ANOVA output). In summary, the effect of the sound source was deemed significant (i.e. $F(2, 166) = 4.32$, $p = 0.015$, $\eta_p^2 = .049$). Examination of the mean values confirmed the male speech source to be significantly different from the two noise sources. This result was expected given the outcome from the previous two tests.

Overall the difference in performance between the decoders was deemed highly significant (i.e. $F(4, 332) = 35.827$, $p < 0.01$, $\eta_p^2 = .302$). This was also the case for listening positions (i.e. $F(4,$

332) = 35.827, $p < 0.01$, $\eta_p^2 = .302$). These findings will be subject of further analysis in the following sections.

When examining the interaction between the variables under test there are a number of interesting features. Firstly, the interaction between sound source and decoder is not deemed significant (i.e. $F(6.50, 539.60) = .775$, $p < 0.599$, $\eta_p^2 = .009$) – a result which contrasts with the previous test for centrally seated listeners. This could be attributed to the fact that on the whole performance is not as good for decoders in off-centre positions (i.e. the influence a sound source has on a decoders performance is less clearly defined when in an off-centre listening position). This hypothesis is further supported by the fact that the interaction between all variables (i.e. sound source, decoder and listening position) is not significant (i.e. $F(15.19, 1260.39) = .775$, $p < 0.909$, $\eta_p^2 = .011$). The other interesting feature is the fact that the interaction between decoder and listening position is highly significant which implies that some decoders perform significantly better for some positions than for others (i.e. $F(8.46, 702.02) = 7.25$, $p < 0.01$, $\eta_p^2 = .011$). This will also be examined in more detail in the following sections.

6.6.3.2 Low frequency noise

Table 6.10 presents the mean localisation error by angle for each decoder for the low frequency noise source (listening positions 1 to 5). Table 6.11 presents the corresponding data for listening positions 6 to 9. The standard deviation of the mean error by angle is also presented in the end column of each table as a measure of the spread of the mean localisation error by angle. To guide the reader better performance is indicated by the lighter shaded cells.

In general, the first order decoder had the highest mean error by angle (as expected). This is due to phantom images being drawn towards the nearest loudspeaker when in off-centre positions because of the significant level of loudspeaker crosstalk for this decoder (note the front-left and left positions in particular in table 6.10).

In order to illustrate the problem of loudspeaker bias figure 6-30 plots the mean user response angle versus the actual source angle for each decoder at the left listening position (P4) and the right listening position (P8). It is clear from this figure that in both positions the mean user response for the first order decoder is biased towards the nearest loudspeaker. In P4 the responses are biased towards the left surround loudspeaker (most obvious when a source is panned to the rear) whereas in P8 the responses are biased towards the right surround loudspeaker (most obvious when panning sources to the front).

The loudspeaker bias was a problem for all decoders. However, the off-centre optimised decoder was able to reduce this problem at all listening positions. Note in figure 6-30 that although loudspeaker bias is still apparent for the fourth order off-centre optimised decoder, it is closer to the ideal response line in both positions particularly around the front of the system.

A 2-way repeated-measures ANOVA (decoder and listening position) revealed the difference between the performance of the decoders to be highly significant, $F(4, 322) = 10.82, p < 0.01, \eta_p^2 = .115$ (see table 14 in Appendix C). This was also the case for listening position $F(2.88, 239.18) = 56.01, p < 0.01, \eta_p^2 = .403$. In order to illustrate this figure 6-31 shows the overall mean error and 95% confidence intervals for all decoders at all listening positions. This figure shows that the fourth order off-centre optimised decoder has the lowest mean error in the majority of listening positions (P1, P2, P3, P4, P5, P6, and P8). The decoder optimised by Poletti has the lowest overall mean error in the back-right position (P7) while the standard fourth order decoder has the lowest overall mean error in the front-right position (P9). The overall mean errors and 95% confidence intervals show that performance of the off-centre optimised decoder was significantly better than the first order decoder in listening positions P2, P3, P5, P6 and P8 and significantly better than the standard fourth order decoder and the fourth decoder by Wiggins at the back listening position (P6). A further interesting aspect that is clarified by figure 6-31 is the significant interaction between decoder and position as flagged by the ANOVA. Some decoders clearly perform better at some position than others. The extreme case being for the first order decoder between the centre and front-left listening positions.

Mean error by angle - low frequency noise																					
Position	Decoder	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	σ
Centre	First order	60.00	31.67	28.57	0.00	50.00	30.00	40.00	10.00	18.33	10.00	10.00	17.14	24.00	86.67	25.00	48.57	103.33	95.00	120.00	35.07
	Standard fourth order	60.00	6.67	8.57	0.00	28.00	10.00	33.33	13.33	21.67	10.00	15.00	18.57	12.00	90.00	30.00	91.43	80.00	90.00	116.67	36.66
	Fourth order off-centre	60.00	30.00	4.29	0.00	36.00	12.50	6.67	15.00	23.33	30.00	5.00	12.86	30.00	30.00	50.00	46.25	66.67	90.00	132.50	33.33
	Fourth order Wiggins	6.67	5.00	31.43	15.00	30.00	12.50	36.67	28.33	16.67	30.00	10.00	15.71	30.00	70.00	45.00	83.75	110.00	85.00	116.67	34.88
	Fourth order Poletti	60.00	55.00	22.86	0.00	44.00	60.00	30.00	11.67	10.00	0.00	20.00	18.57	28.00	33.33	40.00	43.75	23.33	170.00	87.50	38.61
Front	First order	10.00	50.00	21.43	85.00	22.00	20.00	23.33	35.00	15.00	70.00	45.00	52.86	87.50	60.00	130.00	116.25	98.33	90.00	97.50	37.06
	Standard fourth order	6.67	8.33	10.00	5.00	6.00	12.50	16.67	36.67	33.33	40.00	40.00	22.86	60.00	40.00	65.00	104.29	66.67	90.00	85.00	31.12
	Fourth order off-centre	0.00	6.67	14.29	5.00	12.00	15.00	23.33	36.67	26.67	40.00	20.00	35.71	38.00	36.67	57.50	48.75	56.67	90.00	45.00	22.20
	Fourth order Wiggins	10.00	11.67	25.71	15.00	12.00	20.00	26.67	31.67	23.33	60.00	15.00	63.33	56.00	30.00	52.50	101.43	46.67	10.00	110.00	29.68
	Fourth order Poletti	60.00	6.67	12.86	5.00	10.00	7.50	23.33	30.00	31.67	20.00	35.00	38.57	30.00	20.00	20.00	63.75	60.00	10.00	10.00	18.64
Front-left	First order	40.00	50.00	105.00	66.00	85.00	40.00	60.00	26.00	56.67	88.00	100.00	112.50	140.00	110.00	115.00	130.00	152.00	70.00	112.50	36.42
	Standard fourth order	34.00	30.00	40.00	68.00	26.67	34.00	50.00	23.33	23.33	60.00	24.00	80.00	150.00	150.00	100.00	106.67	118.33	100.00	84.00	42.48
	Fourth order off-centre	22.00	26.00	65.00	40.00	60.00	12.00	35.00	30.00	16.67	21.67	26.00	45.00	40.00	96.67	40.00	65.00	101.67	57.50	100.00	27.96
	Fourth order Wiggins	56.00	36.00	77.50	42.00	20.00	28.00	50.00	36.67	36.67	30.00	37.50	62.50	10.00	123.33	97.50	108.33	141.67	75.00	96.00	37.16
	Fourth order Poletti	45.00	18.00	50.00	30.00	70.00	12.50	22.50	35.00	30.00	38.00	32.00	52.50	20.00	96.67	30.00	66.67	100.00	42.50	67.50	25.05
Left	First order	92.00	82.00	75.00	74.00	73.33	88.00	40.75	48.33	10.00	71.67	32.00	55.00	20.00	36.67	117.50	120.00	83.33	77.50	102.50	30.92
	Standard fourth order	88.00	58.00	47.50	76.00	30.00	52.00	42.50	35.00	26.67	45.00	6.00	46.67	130.00	93.33	115.00	113.33	92.00	95.00	72.00	34.15
	Fourth order off-centre	56.00	52.50	62.50	54.00	23.33	44.00	50.00	35.00	16.67	26.67	38.00	15.00	20.00	56.67	57.50	68.33	106.67	92.50	84.00	25.51
	Fourth order Wiggins	94.00	60.00	60.00	58.00	50.00	48.00	85.00	43.33	26.67	36.67	14.00	25.00	20.00	83.33	95.00	101.67	118.33	97.50	82.00	31.20
	Fourth order Poletti	94.00	48.00	52.50	52.50	56.67	40.00	23.33	31.67	20.00	63.33	8.00	33.33	10.00	20.00	22.50	115.00	82.00	90.00	67.50	30.42
Back-left	First order	80.00	112.00	87.50	125.00	95.00	116.00	80.00	56.67	43.33	48.00	14.00	35.00	10.00	80.00	107.50	105.00	75.00	85.00	86.00	33.30
	Standard fourth order	112.00	66.00	102.50	50.00	45.00	67.50	45.00	31.67	40.00	50.00	16.00	20.00	130.00	120.00	100.00	95.00	76.67	80.00	70.00	33.70
	Fourth order off-centre	94.00	60.00	95.00	66.00	60.00	30.00	17.50	25.00	33.33	15.00	14.00	12.50	0.00	50.00	30.00	41.67	95.00	97.50	70.00	31.85
	Fourth order Wiggins	140.00	90.00	80.00	46.67	46.67	56.00	70.00	48.33	36.67	16.67	18.00	10.00	0.00	86.67	95.00	90.00	91.67	85.00	88.00	36.27
	Fourth order Poletti	116.00	32.00	130.00	74.00	86.67	44.00	12.50	45.00	43.33	58.33	28.00	45.00	10.00	16.67	32.50	46.67	93.33	90.00	64.00	34.21

Table 6.10: Mean error by angle for the low frequency noise source - positions 1 to 5

Mean error by angle - low frequency noise																					
Position	Decoder	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	σ
Back	First order	6.67	66.00	98.33	140.00	106.00	130.00	56.67	46.67	31.67	20.00	65.00	24.29	42.00	23.33	95.00	68.75	100.00	90.00	80.00	38.25
	Standard fourth order	60.00	8.33	65.71	60.00	92.00	85.00	73.33	40.00	38.33	20.00	25.00	12.86	56.00	26.67	70.00	86.25	100.00	10.00	115.00	32.50
	Fourth order off-centre	63.33	5.00	28.57	5.00	68.00	27.50	56.67	38.33	28.33	20.00	10.00	5.71	16.00	10.00	25.00	38.75	66.67	15.00	80.00	24.01
	Fourth order Wiggins	63.33	66.67	54.29	80.00	82.00	55.00	53.33	38.33	28.33	20.00	15.00	2.86	34.00	30.00	45.00	102.50	101.67	95.00	120.00	32.85
	Fourth order Poletti	120.00	33.33	45.71	5.00	66.00	45.00	46.67	43.33	35.00	20.00	15.00	5.71	18.00	30.00	32.50	38.75	71.67	95.00	122.50	34.38
Back-right	First order	63.33	122.00	107.14	110.00	70.00	52.50	40.00	36.67	21.67	20.00	10.00	7.14	18.00	23.33	30.00	40.00	50.00	60.00	72.50	34.07
	Standard fourth order	3.33	31.67	82.86	120.00	76.00	57.50	43.33	38.33	28.33	20.00	10.00	5.71	10.00	26.67	30.00	40.00	50.00	60.00	97.50	32.14
	Fourth order off-centre	10.00	35.00	72.86	65.00	97.50	62.50	36.67	30.00	28.33	10.00	10.00	5.71	14.00	20.00	32.50	42.50	50.00	65.00	85.00	27.43
	Fourth order Wiggins	20.00	56.67	72.86	125.00	66.00	50.00	40.00	38.33	31.67	30.00	10.00	5.71	10.00	23.33	32.50	45.00	55.00	60.00	75.00	28.59
	Fourth order Poletti	13.33	10.00	68.57	65.00	32.00	55.00	43.33	35.00	25.00	30.00	5.00	7.14	14.00	23.33	30.00	43.75	51.67	70.00	75.00	22.52
Right	First order	50.00	73.33	100.00	65.00	40.00	42.50	33.33	26.67	21.67	10.00	5.00	11.43	26.00	26.67	52.50	58.75	68.33	75.00	85.00	27.11
	Standard fourth order	3.33	38.33	42.86	60.00	44.00	35.00	23.33	15.00	21.67	10.00	15.00	14.29	30.00	23.33	37.50	52.50	51.67	75.00	72.50	20.69
	Fourth order off-centre	13.33	38.33	22.86	10.00	26.00	30.00	20.00	20.00	0.00	15.00	14.29	28.00	26.67	37.50	50.00	66.67	60.00	82.50	20.86	
	Fourth order Wiggins	13.33	55.00	58.57	60.00	70.00	47.50	23.33	16.67	21.67	10.00	20.00	8.57	26.00	23.33	52.50	50.00	61.67	70.00	87.50	24.07
	Fourth order Poletti	20.00	33.33	75.71	10.00	30.00	17.50	33.33	31.67	28.33	10.00	20.00	11.43	26.00	20.00	40.00	46.25	63.33	75.00	85.00	23.16
Front-right	First order	36.67	26.67	44.29	10.00	32.00	17.50	46.67	45.00	55.00	50.00	55.00	64.29	46.00	80.00	67.50	62.50	90.00	95.00	70.00	22.76
	Standard fourth order	93.33	60.00	22.86	0.00	10.00	20.00	33.33	43.33	53.33	50.00	60.00	25.71	12.00	33.33	52.50	51.25	65.00	75.00	75.00	25.01
	Fourth order off-centre	3.33	38.00	68.33	0.00	10.00	20.00	30.00	41.67	60.00	40.00	20.00	34.29	34.00	53.33	36.67	54.29	66.00	115.00	53.33	26.75
	Fourth order Wiggins	76.67	44.00	26.67	0.00	10.00	20.00	30.00	43.33	53.33	50.00	50.00	34.29	14.00	73.33	56.67	65.71	68.00	105.00	40.00	25.99
	Fourth order Poletti	6.67	92.00	68.33	5.00	10.00	20.00	33.33	40.00	58.33	40.00	5.00	45.71	32.00	60.00	60.00	64.29	82.00	130.00	53.33	32.62

Table 6.11: Mean error by angle for the low frequency noise source - position 6 to 9

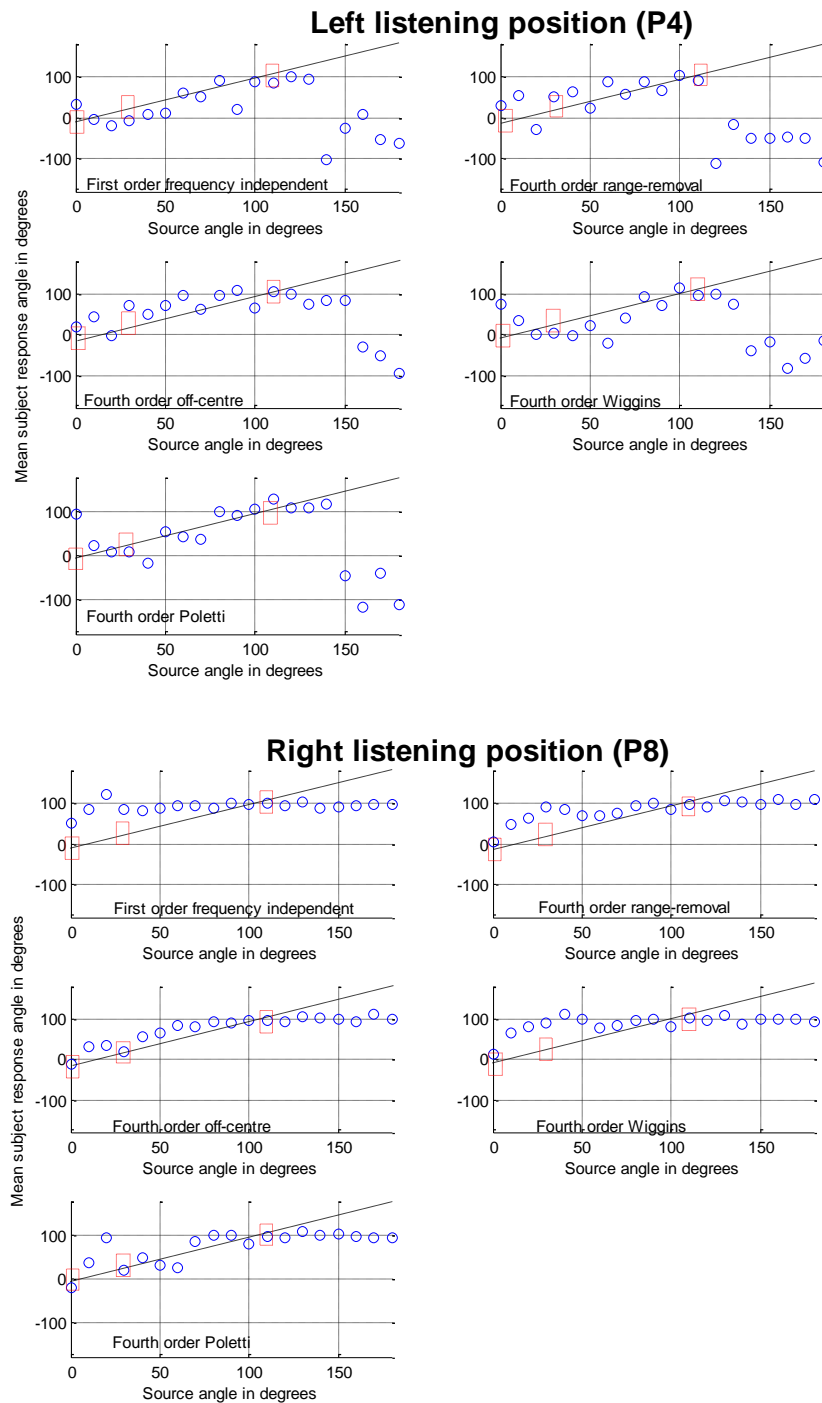


Figure 6-30: Mean response angle versus the actual source angle for each decoder in the low frequency noise test (listening positions 4 and 8). Loudspeakers are shown as red squares. Dashed line is the ideal response. Note the loudspeaker bias effect, particularly in position 4.

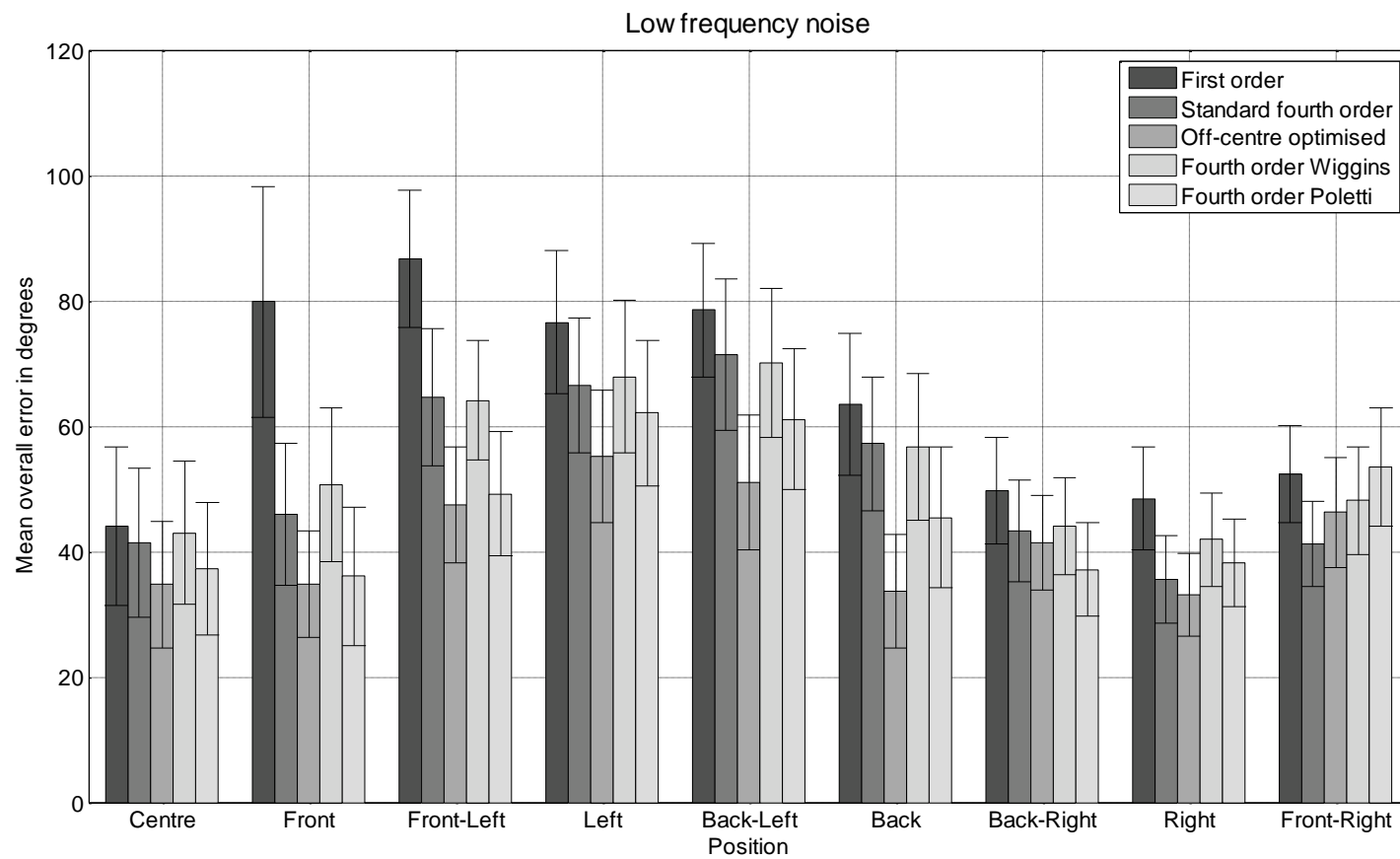


Figure 6-31: Overall mean errors with 95% confidence intervals for each decoder at each listening position for the low frequency noise source.

6.6.3.3 Mid/high frequency noise

Table 6.12 and table 6.13 present the mean localisation error by angle for each decoder for the mid/high frequency noise source. The standard deviation (presented in the end column of each table) describes the spread of the error for each decoder at each position. To guide the reader better performance is indicated by the lighter shaded cells.

In almost all off-centre listening positions the loudspeaker nearfield bias effect was a problem. However, selected results presented in figure 6-32 demonstrate that the off-centre optimised decoder and the decoder developed by Poletti were able to reduce this effect when compared with the other decoders under test. For the front listening position (P2) when a source is panned to the rear the mean subject response is much closer to the ideal response angle. This is also the case for the rear position (P6) when source is panned to the front (particularly for Poletti's decoder).

Examination of the mean error values by angle reveals that the fourth order off-centre optimised decoder and the fourth order decoder by Poletti generally gave better performance by angle than the other decoders. The biggest difference at all listening positions appears to be when presenting sound sources in the front of the listener and the side of the listener.

A 2-way repeated-measures ANOVA test was undertaken to determine the level of significance for decoder and listening position (see table 15 for the full results). The ANOVA flagged a significant difference in terms for decoder performance $F(4, 322) = 12.12, p < 0.01, \eta_p^2 = .127$ and position $F(2.585, 214.58) = 60.04, p < 0.01, \eta_p^2 = .420$. Figure 6-33 displays the overall mean error for each decoder at each listening position. The data shows that the off-centre optimised decoder performed significantly better than first order decoder at P2, P3, P4, P5, P6 and P9. In addition, the off-centre optimised decoder performed significantly better than the standard fourth order decoder and the decoder optimised by Wiggins at P3, P4, P5, and P6.

All fourth order decoders were significantly better than the first order decoder at P2, P3, P4, and P9 and the Poletti decoder was significantly better than the standard fourth order decoder in P4.

Mean error by angle - mid/high frequency noise																					
Position	Decoder	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	σ
Centre	First order	0.00	12.50	20.00	48.00	10.00	30.00	50.00	37.50	28.75	20.00	35.00	33.33	62.00	110.00	72.50	70.00	47.50	10.00	110.00	31.27
	Standard fourth order	0.00	30.00	8.33	8.00	13.33	20.00	17.50	20.00	18.75	28.00	21.67	23.33	28.33	107.50	92.50	70.00	87.50	10.00	160.00	42.72
	Fourth order off-centre	0.00	45.00	3.33	2.00	13.33	23.33	15.00	12.50	27.50	18.00	20.00	11.67	26.67	35.00	25.00	42.50	47.50	20.00	123.33	27.10
	Fourth order Wiggins	0.00	5.00	11.67	12.00	10.00	20.00	22.50	15.00	26.25	22.00	25.00	20.00	15.00	70.00	100.00	30.00	47.50	10.00	111.67	30.97
	Fourth order Poletti	0.00	85.00	1.67	2.00	10.00	20.00	22.50	37.50	23.75	18.00	20.00	15.00	20.00	37.50	30.00	45.00	40.00	10.00	140.00	33.01
Front	First order	10.00	10.00	15.00	18.00	20.00	43.33	27.50	35.00	45.00	48.00	64.00	113.33	117.50	125.00	160.00	137.50	115.00	10.00	122.00	51.64
	Standard fourth order	0.00	10.00	18.33	6.00	13.33	20.00	27.50	37.50	46.25	44.00	36.67	40.00	46.00	92.50	92.50	27.50	65.00	10.00	84.00	28.42
	Fourth order off-centre	0.00	10.00	18.33	2.50	3.33	20.00	30.00	42.50	42.50	36.00	21.67	35.00	33.33	37.50	22.50	27.50	50.00	10.00	58.00	16.51
	Fourth order Wiggins	0.00	5.00	13.33	8.00	10.00	20.00	22.50	42.50	42.50	54.00	43.33	54.00	43.33	80.00	87.50	82.50	60.00	170.00	72.00	40.41
	Fourth order Poletti	0.00	12.50	36.67	4.00	6.67	23.33	32.50	45.00	42.50	42.00	44.00	35.00	31.67	65.00	40.00	60.00	72.50	10.00	118.00	28.13
Front-left	First order	30.00	40.00	66.67	70.00	96.00	62.50	26.67	48.00	85.00	110.00	118.00	126.67	150.00	140.00	140.00	130.00	162.50	117.50	136.67	43.10
	Standard fourth order	38.57	15.00	13.33	12.50	11.67	15.00	23.33	36.00	25.00	50.00	68.00	80.00	92.50	110.00	120.00	110.00	135.00	112.50	100.00	43.83
	Fourth order off-centre	1.43	10.00	13.33	22.50	15.00	12.50	16.67	30.00	15.00	46.00	34.00	38.33	32.50	44.00	90.00	110.00	96.67	95.00	101.67	36.07
	Fourth order Wiggins	32.86	38.33	30.00	12.50	11.67	17.50	16.67	36.00	35.00	48.00	46.00	48.33	100.00	134.00	116.67	160.00	100.00	115.00	86.67	45.69
	Fourth order Poletti	4.29	8.33	16.67	7.50	5.00	17.50	20.00	30.00	45.00	58.00	60.00	56.67	42.50	64.00	126.67	120.00	93.33	75.00	86.67	38.17
Left	First order	18.57	38.00	50.00	70.00	86.67	77.50	83.33	84.00	100.00	60.00	72.00	102.00	115.00	116.00	126.67	115.00	100.00	87.50	80.00	28.03
	Standard fourth order	28.57	13.33	13.33	2.50	10.00	22.50	23.33	38.00	80.00	46.00	40.00	108.00	105.00	122.00	116.67	100.00	100.00	85.00	84.17	41.62
	Fourth order off-centre	2.86	5.00	3.33	5.00	8.33	17.50	40.00	60.00	40.00	32.00	14.00	16.67	30.00	24.00	50.00	45.00	106.67	87.50	71.67	29.83
	Fourth order Wiggins	22.86	23.33	6.67	5.00	6.67	17.50	26.67	62.00	25.00	40.00	34.00	31.67	112.50	117.50	123.33	110.00	95.00	97.50	79.17	42.47
	Fourth order Poletti	2.86	8.33	0.00	0.00	10.00	25.00	33.33	50.00	50.00	38.00	20.00	25.00	32.50	52.00	40.00	72.50	83.33	95.00	81.67	29.24
Back-left	First order	42.86	43.33	45.00	57.50	67.50	92.50	27.50	72.50	40.00	50.00	20.00	71.67	77.50	140.00	110.00	100.00	82.50	75.00	73.33	29.74
	Standard fourth order	6.43	10.00	6.67	47.50	10.00	20.00	25.00	38.00	35.00	42.00	16.00	72.00	112.50	128.00	110.00	70.00	76.67	80.00	70.00	38.40
	Fourth order off-centre	5.71	8.33	16.67	2.50	10.00	17.50	68.33	38.00	30.00	15.00	12.00	8.33	7.50	16.00	30.00	25.00	57.50	80.00	70.00	24.20
	Fourth order Wiggins	15.71	28.33	6.67	7.50	11.67	10.00	15.00	46.00	30.00	14.00	14.00	41.67	70.00	130.00	93.33	100.00	90.00	80.00	71.67	38.41
	Fourth order Poletti	6.43	5.00	6.67	0.00	11.67	22.50	26.67	68.00	35.00	52.00	28.00	20.00	12.50	40.00	56.67	30.00	80.00	77.50	68.33	25.95

Table 6.12: Mean error by angle for the mid/high frequency noise source - positions 1 to 5

Mean error by angle - mid/high frequency noise																					
Position	Decoder	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	σ
Back	First order	10.00	82.50	21.67	84.00	73.33	45.00	40.00	25.00	26.25	18.00	15.00	23.33	33.33	30.00	62.50	97.50	37.50	160.00	42.00	37.13
	Standard fourth order	0.00	7.50	3.33	4.00	46.67	60.00	42.50	32.50	35.00	26.00	10.00	15.00	36.67	27.50	40.00	60.00	25.00	170.00	66.00	37.84
	Fourth order off-centre	0.00	45.00	6.67	2.00	46.67	36.67	40.00	37.50	28.75	26.00	11.67	6.67	13.33	17.50	32.50	37.50	35.00	10.00	91.67	21.66
	Fourth order Wiggins	10.00	10.00	15.00	24.00	43.33	50.00	42.50	35.00	30.00	14.00	10.00	13.33	28.33	22.50	35.00	75.00	50.00	170.00	68.00	37.03
	Fourth order Poletti	10.00	42.50	3.33	0.00	6.67	13.33	26.67	17.50	23.75	16.00	13.33	6.67	10.00	12.50	35.00	57.50	35.00	170.00	40.00	37.55
Back-right	First order		22.50	71.67	48.00	73.33	56.67	37.50	35.00	27.50	18.00	10.00	3.33	8.33	17.50	32.50	42.50	50.00	70.00	70.00	23.15
	Standard fourth order	10.00	7.50	16.67	30.00	60.00	60.00	40.00	37.50	28.75	18.00	10.00	3.33	6.67	20.00	30.00	42.50	50.00	60.00	68.33	20.85
	Fourth order off-centre	0.00	25.00	24.00	2.00	23.33	56.67	35.00	40.00	30.00	20.00	10.00	1.67	10.00	17.50	30.00	40.00	52.50	60.00	71.67	20.69
	Fourth order Wiggins		20.00	14.00	52.50	66.67	63.33	40.00	37.50	30.00	16.00	10.00	1.67	6.67	20.00	30.00	42.50	55.00	60.00	75.00	22.57
	Fourth order Poletti	0.00	12.50	26.67	4.00	20.00	40.00	53.33	35.00	26.25	16.00	8.33	1.67	8.33	20.00	32.50	40.00	47.50	60.00	68.33	20.23
Right	First order	40.00	27.50	10.00	6.00	10.00	20.00	30.00	40.00	46.25	65.00	55.00	45.00	46.67	80.00	53.33	16.67	70.00		93.33	24.94
	Standard fourth order	10.00	20.00	10.00	0.00	10.00	23.33	27.50	40.00	48.75	52.00	48.33	18.33	21.67	57.50	50.00	32.50	45.00	40.00	73.33	19.60
	Fourth order off-centre	10.00	5.00	8.33	2.00	10.00	20.00	30.00	40.00	35.00	42.00	23.33	15.00	23.33	77.50	62.50	30.00	65.00	40.00	80.00	24.01
	Fourth order Wiggins	30.00	17.50	10.00	0.00	10.00	23.33	27.50	40.00	46.25	54.00	53.33	24.00	20.00	37.50	52.50	27.50	40.00	30.00	71.67	17.99
	Fourth order Poletti	0.00	27.50	8.33	0.00	10.00	20.00	32.50	42.50	52.50	44.00	21.67	18.33	20.00	55.00	45.00	32.50	40.00	50.00	70.00	19.55
Front-right	First order	40.00	27.50	10.00	6.00	10.00	20.00	30.00	40.00	46.25	65.00	55.00	45.00	46.67	80.00	53.33	16.67	70.00		93.33	24.94
	Standard fourth order	10.00	20.00	10.00	0.00	10.00	23.33	27.50	40.00	48.75	52.00	48.33	18.33	21.67	57.50	50.00	32.50	45.00	40.00	73.33	19.60
	Fourth order off-centre	10.00	5.00	8.33	2.00	10.00	20.00	30.00	40.00	35.00	42.00	23.33	15.00	23.33	77.50	62.50	30.00	65.00	40.00	80.00	24.01
	Fourth order Wiggins	30.00	17.50	10.00	0.00	10.00	23.33	27.50	40.00	46.25	54.00	53.33	24.00	20.00	37.50	52.50	27.50	40.00	30.00	71.67	17.99
	Fourth order Poletti	0.00	27.50	8.33	0.00	10.00	20.00	32.50	42.50	52.50	44.00	21.67	18.33	20.00	55.00	45.00	32.50	40.00	50.00	70.00	19.55

Table 6.13: Mean error by angle for the mid/high frequency noise source - positions 6 to 9

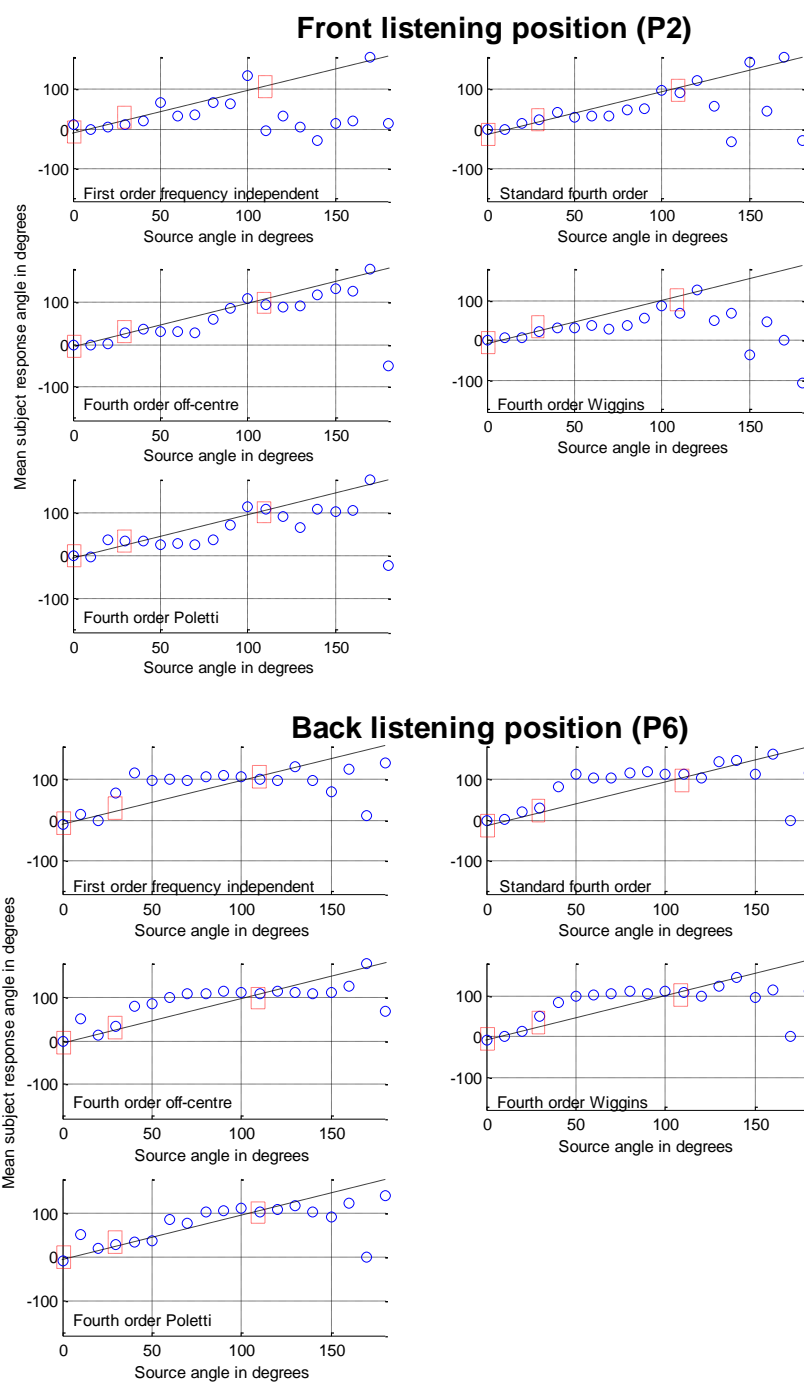


Figure 6-32: Mean response angle versus the actual source angle for each decoder in the mid/high frequency noise test (listening positions 2 and 6). Loudspeakers are shown as red squares. Dashed line is the ideal response. Note the loudspeaker bias effect in both plots.

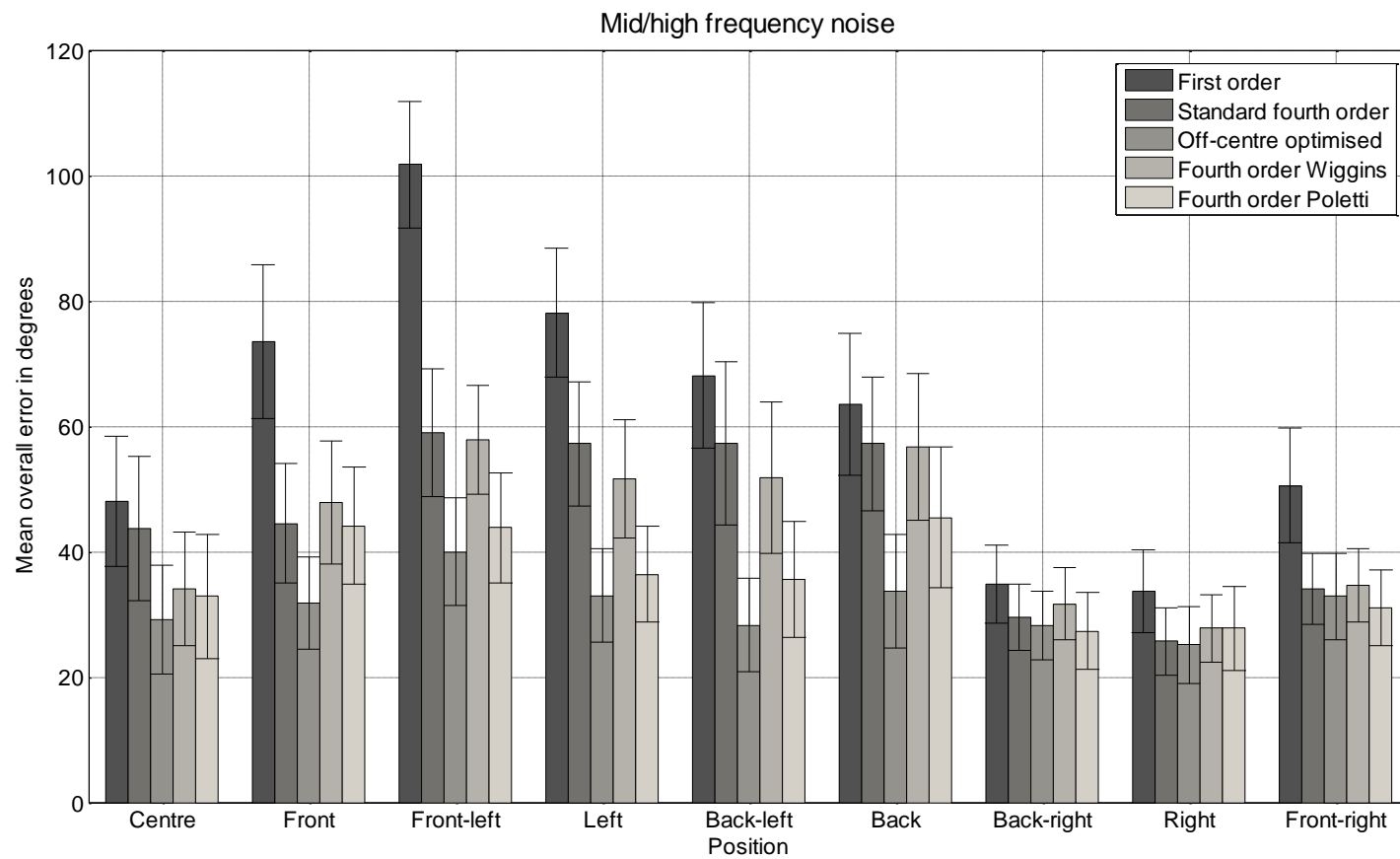


Figure 6-33: Overall mean errors with 95% confidence intervals for each decoder at each listening position for the mid/high frequency noise source.

6.6.3.4 Male speech

Table 6.14 and table 6.15 present the mean localisation error by angle for the male speech test. The standard deviation of the mean error by angle is presented in the end column of each table. In order to guide the reader better performance is indicated by the lighter shaded cells.

In this test loudspeaker bias effects were apparent for all decoders in off-centre listening positions. However, selected plots displayed in figure 6-34 demonstrate that the off-centre optimised decoder and the decoder developed by Poletti were able to considerably reduce this problem. For example, note that at the front listening position (P2) sources to the rear are much closer to the ideal response line for both decoders. This is also the case for the front-left listening position (P3).

A 2-way repeated measures ANOVA revealed that the differences between the performance of the decoders and also the listening position was highly significant (i.e. $F(4, 322) = 14.24, p < 0.01, \eta_p^2 = .146$ and $F(2.26, 187.35) = 82.98, p < 0.01, \eta_p^2 = .500$ respectively) (see table 16 of Appendix C for full statistical details). To illustrate this figure 6-35 shows the Poletti decoder had the lowest overall mean error for 6 out of 9 positions (P2, P3, P4, P7, P8 and P9) whereas the fourth order off-centre optimised decoder had the lowest overall mean error at two positions (P5 and P6). Please note that the error values for both off-centre optimised decoders would have been much lower at the centre position if the front-back reversals at 170° were removed (see table 6.14).

The overall mean errors and 95% confidence intervals in figure 6-35 shows that the two off-centre optimised decoders gave significantly better performance than the other decoders at all left side listening positions (P3, P4 and P5). The off-centre decoder produced by the design tool also gave significantly better performance than the first order decoder at the rear listening position (P6). In addition, all fourth order decoders were deemed to be significantly better than the first order decoder in all front, left-front and left listening position (P2, P3 and P4).

Mean error by angle - male speech																					
Position	Decoder	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	σ
Centre	First order	0	10	12.5	12.5	10	20	20	13.333		28	7.5	5	10	24	32.857	13.333	17.143	10	10	8.18
	Standard fourth order	0	6	10	2.5	6.6667	12.5	0	13.333		12	10	12.5	18.333	20	42.857	23.333	25.714	10	8.3333	10.36
	Fourth order off-centre	3.3333	8	7.5	2.5	10	20	10	16.667		14	10	5	6.6667	18	31.429	26.667	24.286	140	31.667	30.99
	Fourth order Wiggins	3.3333	10	12.5	2.5	6.6667	10	10	10		26	15	5	11.667	12	41.429	26.667	11.429	25	30	10.52
	Fourth order Poletti	0	6	5	0	6.6667	22.5	25	20		14	10	5	10	22	25.714	16.667	15.714	90	31.667	20.19
Front	First order	0	40	30	22.5	23.333	25	30	40		50	47.5	72.5	86.667	148	148.57	180	71.429	45	116.67	51.07
	Standard fourth order	0	26	30	12.5	10	20	25	30		40	15	30	40	58	71.429	96.667	50	50	58.333	24.11
	Fourth order off-centre	0	26	35	5	10	20	30	26.667		24	47.5	0	13.333	20	28.571	36.667	47.143	35	97	22.32
	Fourth order Wiggins	0	12	27.5	15	16.667	20	30	50		46	12.5	10	15	26	65.714	80	64.286	40	48.333	22.54
	Fourth order Poletti	0	26	7.5	27.5	10	20	35	43.333		22	32.5	30	11.667	32	31.429	33.333	28.571	35	68.333	15.21
Front-left	First order	20	40	85	70	135	12.5	95	40	30	10	10	7.5	32	132.5	114.44	102.5	90	80	70	42.84
	Standard fourth order	7.5	5	5	8.3333	10	7.5	30	20	26.667	20	12.5	12.5	80	125	110	102.5	90	80	70	41.82
	Fourth order off-centre	2.5	6.25	8.3333	8.3333	10	7.5	24.286	40	30	18	7.5	0	10	22.5	36.667	37.5	100	60	70	26.40
	Fourth order Wiggins	17.5	17.5	13.333	9.1667	10	27.5	35.714	40	30	16	5	2.5	53.333	125	110	102.5	90	80	70	39.29
	Fourth order Poletti	2.5	5	6.6667	1.6667	10	20	22.857	10	20	44	12.5	5	8.3333	20	24.444	60	95	80	70	28.47
Left	First order	30	40	70	105	150	67.5	28.333	40	23.333	54	22.5	15	86.667	112.5	112.5	100	90	80	71.667	37.80
	Standard fourth order	25	15	6.6667	6.6667	10	12.5	24.286	30	23.333	32	20	30	83.333	117.5	110	105	95	80	73.333	39.18
	Fourth order off-centre	2.5	10	11.667	16.667	15	20	28.571	40	30	16	10	10	15	20	34.444	32.5	85	83.333	72.5	24.84
	Fourth order Wiggins	30	27.5	16.667	6.6667	10	17.5	24.286	20	26.667	14	15	7.5	25	90	106.67	102.5	90	73.333	73.333	35.19
	Fourth order Poletti	0	7.5	1.6667	13.333	5	17.5	30	40	50	14	15	0	13.333	20	26.25	37.5	90	73.333	70	26.42
Back-left	First order	30	40	50	60	60	72.5	30	30	46.667	70	130	110	143.33	167.5	151.11	170	115	113.33	85	47.96
	Standard fourth order	22.5	27.5	20	19.167	10	20	25.714	30	46.667	38	30	20	88.333	125	108.89	100	90	86.667	75	36.99
	Fourth order off-centre	2.5	10	21.667	28.333	15	17.5	30	20	36.667	18	10	5	11.667	45	27.778	27.5	65	80	71.667	22.26
	Fourth order Wiggins	30	37.5	56.667	23.333	10	20	24.286	30	53.333	68	36.25	2.5	8.3333	117.5	103.33	102.5	90	86.667	71.667	35.70
	Fourth order Poletti	0	10	16.667	3.3333	10	20	31.429	30	46.667	40	17.5	10	10	23.75	23.333	20	60	80	73.333	22.83

Table 6.14: Mean error by angle for the male speech source – positions 1 to 5

Mean error by angle - male speech																					
Position	Decoder	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	σ
Back	First order	13.333	32	77.5	77.5	70	35	40	40		18	7.5	15	30	40	52.857	10	72.857	155	53.333	35.59
	Standard fourth order	40	6	12.5	5	3.3333	25	40	43.333		24	10	10	36.667	38	72.857	36.667	64.286	90	35	24.37
	Fourth order off-centre	53.333	6	10	5	6.6667	30	25	40		34	10	10	11.667	22	32.857	40	42.857	75	21.667	19.05
	Fourth order Wiggins	46.667	8	15	2.5	46.667	32.5	40	36.667		24	10	2.5	23.333	40	32.857	56.667	97.143	155	40	36.67
	Fourth order Poletti	40	6	2.5	35	10	12.5	10	30		26	5	2.5	8.3333	24	44.286	43.333	64.286	75	43.333	21.75
Back-right	First order	30	20	10	0	13.333	20	30	36.667		56	52.5	15	38.333	6	20	33.333	37.143	30	60	16.95
	Standard fourth order	26.667	16	10	0	10	20	30	43.333		58	30	10	3.3333	20	32.857	33.333	50	50	60	18.51
	Fourth order off-centre	0	8	5	0	10	20	30	40		40	12.5	5	15	26	28.571	33.333	55.714	60	63.333	20.35
	Fourth order Wiggins	30	20	10	0	10	20	30	40		60	37.5	12.5	8.3333	24	27.143	36.667	51.429	65	63.333	19.72
	Fourth order Poletti	3.3333	6	10	0	10	20	30	40		40	17.5	7.5	8.3333	16	25.714	36.667	47.143	50	51.667	17.22
Right	First order	53.333	22	10	52.5	60	45	35	33.333		14	5	2.5	13.333	26	32.857	43.333	50	70	73.333	21.83
	Standard fourth order	16.667	14	10	2.5	10	5	25	30		16	10	2.5	13.333	24	31.429	46.667	50	55	75	20.36
	Fourth order off-centre	0	10	2.5	5	13.333	7.5	35	33.333		20	5	0	13.333	26	31.429	40	51.429	60	70	21.33
	Fourth order Wiggins	23.333	20	10	12.5	16.667	25	40	30		20	5	0	11.667	22	32.857	40	48.571	65	68.333	19.09
	Fourth order Poletti	0	10	5	5	10	20	25	23.333		22	12.5	0	10	22	34.286	40	52.857	65	61.667	20.15
Front-right	First order	23.333	18	80	70	63.333	47.5	35	33.333		18	10	0	11.667	20	30	43.333	50	65	68.333	23.88
	Standard fourth order	6.6667	10	5	5	30	30	35	36.667		20	10	0	10	20	32.857	40	51.429	65	73.333	21.20
	Fourth order off-centre	3.3333	6	7.5	0	10	32.5	35	36.667		20	10	0	10	20	31.429	40	50	60	71.667	21.28
	Fourth order Wiggins	16.667	20	10	60	63.333	45	40	36.667		20	10	0	10	20	32.857	40	51.429	60	71.667	21.57
	Fourth order Poletti	3.3333	2	2.5	5	10	10	10	30		18	10	2.5	10	20	31.429	40	50	60	70	21.06

Table 6.15: Mean error by angle for the male speech source - positions 6 to 9

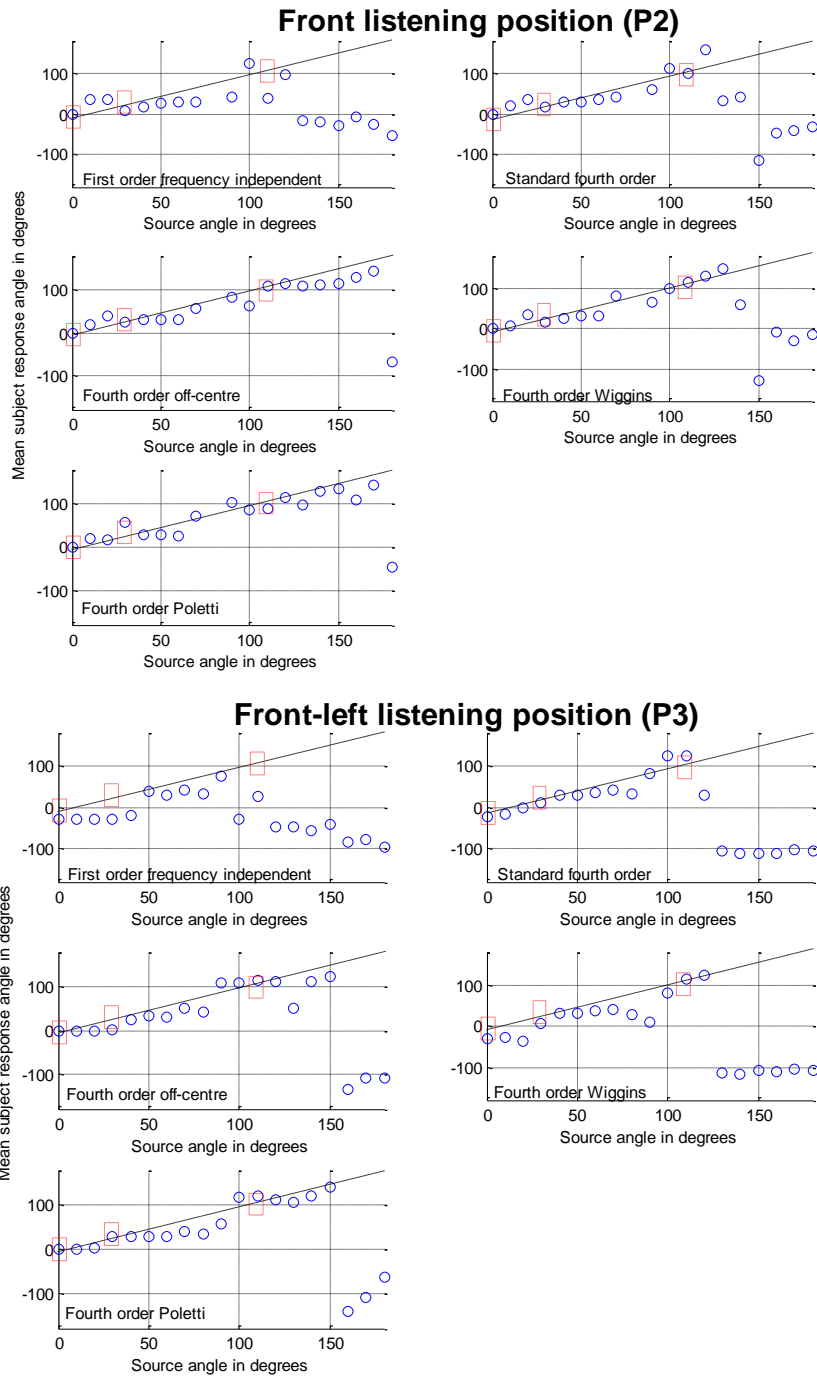


Figure 6-34: Mean response angle versus the actual source angle for each decoder in the male speech test (listening positions 2 and 7). Loudspeakers are shown as red squares. Dashed line is the ideal response. Note the loudspeaker bias effect in both plots.

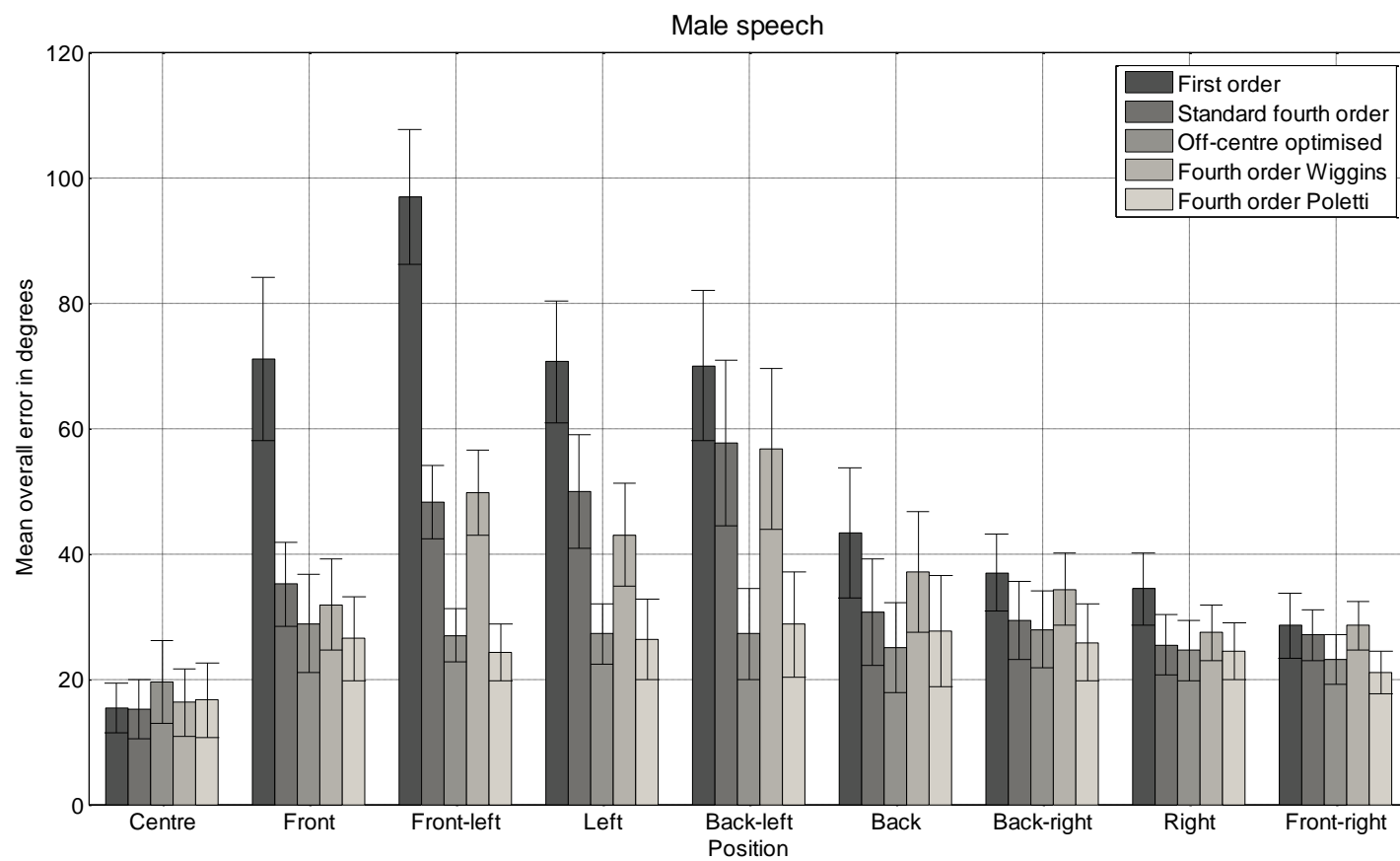


Figure 6-35 : Overall mean errors with 95% confidence intervals for each decoder at each listening position for the male speech source.

6.6.3.5 Overall decoder performance

An overall localisation error was derived for all decoders at each listening position by averaging the results from each sound source test (see figure 6-37). The low error values suggest that the off-centre optimised decoder produced by the design tool had the best localisation performance in 7 out of 9 listening positions (P1, P2, P3, P4, P5, P6, and P8). At the back-right listening position (P7) the low error value suggests that the decoder by Poletti performed the best, whereas at the front-right listening position (P9) the standard fourth order decoder appears to perform the best (note the difference between the standard fourth order decoder and the off-centre optimised decoder at P9 is marginal).

The data shows that at the left side listening positions (P3, P4 and P5) the off-centre optimised decoder and the decoder by Poletti gave significantly better performance than the other decoders. In addition, the off-centre optimised decoder was significantly better than the non off-centre optimised decoders in the front and rear listening positions (P2 and P6).

As expected, the first order decoder performed the worst overall in all listening positions. The performance of the standard fourth order decoder and the fourth order decoder by Wiggins was comparable in most listening positions.

Finally, an overall mean was produced taking into account all listening positions and all sound sources (see figure 6-37). The figure shows that overall the off-centre optimised decoder and the decoder by Poletti were significantly better than first order decoder, standard fourth order decoder and the fourth order decoder by Wiggins. Throughout this analysis the off-centre optimised decoder and the decoder produced by Poletti were very evenly matched and there were never any significant differences between the two.

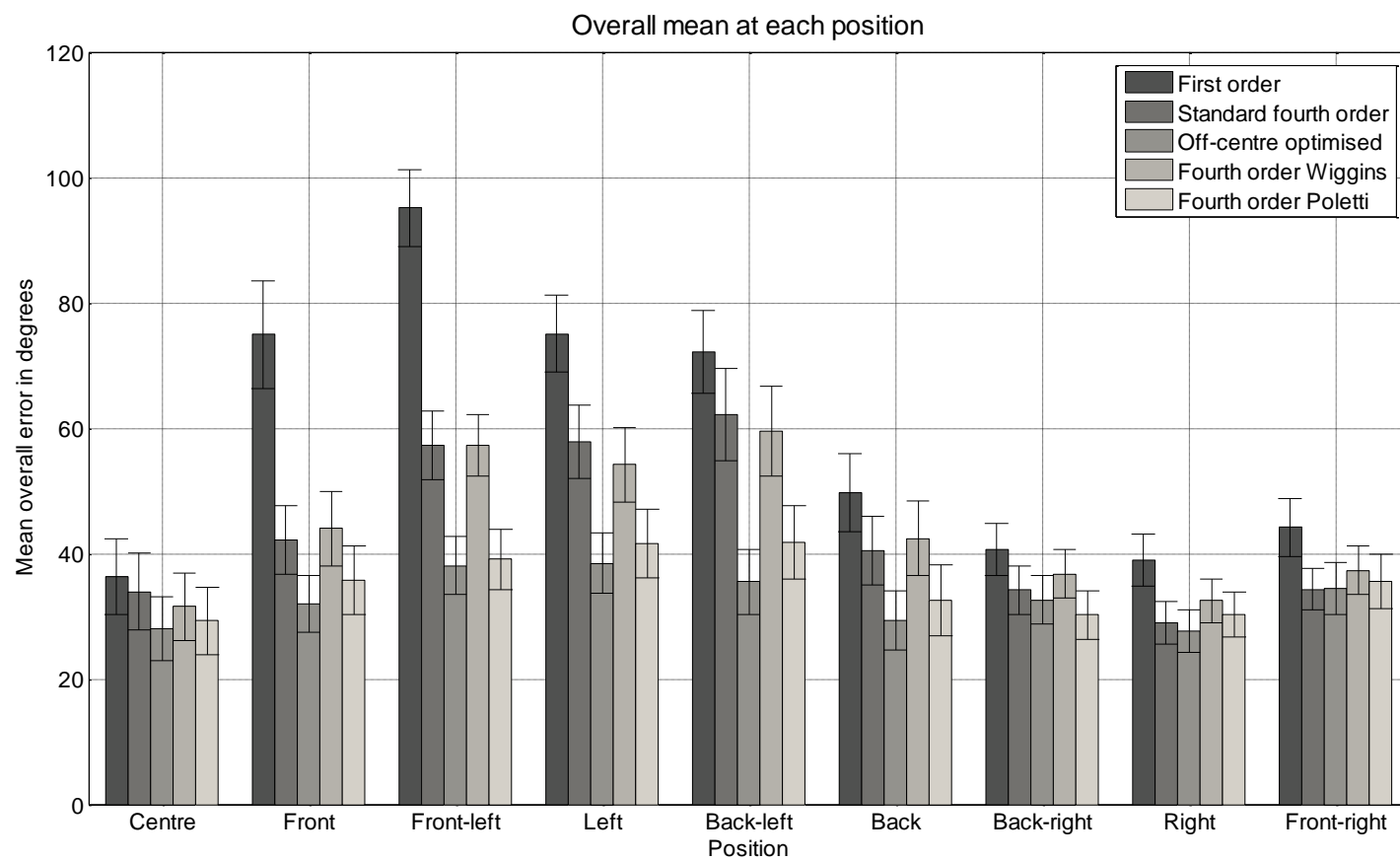


Figure 6-36: Overall mean localisation error at each position (with 95% confidence intervals) taking into account the data from all the sound source tests.

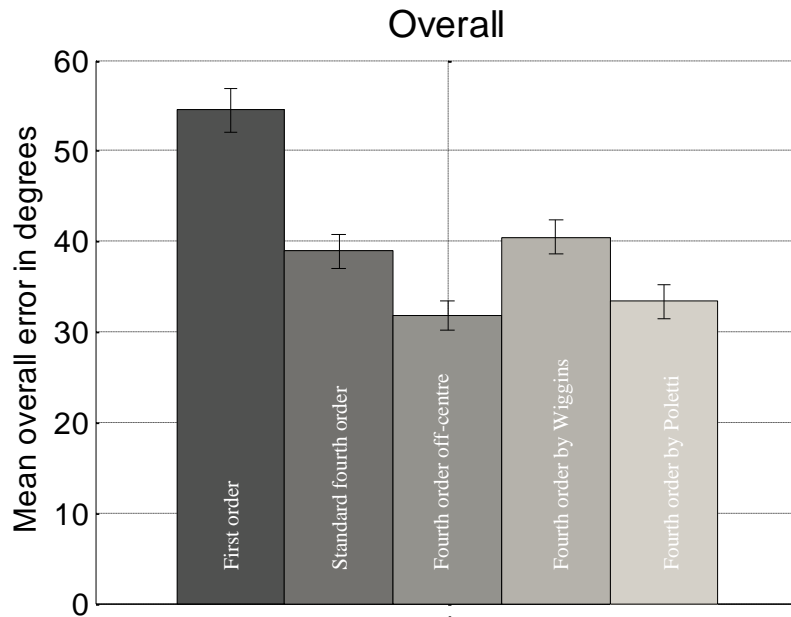


Figure 6-37: Overall mean error and 95% confidence intervals for each decoder taking into account all positions for each sound source test.

6.6.4 Discussion

The localisation performance of five different Ambisonic decoders was evaluated in an off-centre listening test. It was found that in most off-centre listening positions loudspeaker bias was a problem because of the close proximity of the loudspeakers to the listeners (irrespective of decoder and stimuli). Listeners in positions P2, P4, P6 and P8 were situated at 35% of the total surround sound rig radius, whereas listeners in positions P3, P5, P7 and P9 were situated at 50% of the total radius.

In worse cases phantom sound sources were perceived by listeners to come from different directions and some listeners found it difficult to say exactly where certain sounds were coming from (especially for the noise bursts). During informal questioning after the test some listeners commented that the noise bursts were “diffuse” or “spread out” appearing to come from two separate locations on occasions. The problem of localising noisy signals reproduced by Ambisonic systems has been noted in previous research (Gerzon 1985; Benjamin et al. 2006) and

it is likely to be due to the angular spacing between the loudspeakers. Loudspeaker angular spacing changes with listening position giving each listener a different perspective of the sound source image. For example, at the right side listening position the angle between the front-right loudspeaker and back-right loudspeaker is wider than the front-left and back-left position.

Despite these problems, however, there were clear differences between the performances of the decoders proving that it is possible to improve off-centre localisation performance. The fourth order off-centre optimised decoder and the fourth order decoder developed by Poletti were able to significantly reduce loudspeaker bias effects in most off-centre listening positions. The greatest improvement for these decoders was made for listening positions on the opposite side of the system to the sound source (highlighted by the significantly lower mean errors in listening positions P3, P4 and P5). This would undoubtedly be an important factor when reproducing sources to multiple listeners in any listening situation.

Overall, the results presented in this test indicate the method incorporated into the design tool for improving the localisation performance of a decoder in off-centre position was successful. Furthermore, the listening tests establish that the theoretical design criteria used both in this work, and the work of Poletti's, are able to produce decoders for irregular 5-speaker systems with improved off-centre performance when compared to existing decoders.

6.7 Summary

This chapter described a series of listening tests designed to assess the localisation performance of 5-speaker Ambisonic decoders. Listeners were required to localise sound sources at different angles in the horizontal plane. In the first test, listeners were asked to localise real sources of sound so that a direct comparison could be made with decoded sources panned to the same location.

In the second test, the localisation performance of the decoders was evaluated from the central listening position. The results from this test correlated reasonably well with the predictions of the

velocity vector and energy vector. This was especially true for the performance of the MAA optimised decoder which gave good frontal performance for all sound sources. On balance, this was also true for the even error optimised decoder. Even performance was observed for the male speech source in particular.

In some cases, however, the theoretical predictions were not always synonymous with the listening test results (e.g. the velocity vector response of the first order decoders). The large number of front-back reversals subjects encountered during the test did not help in this regard. The next chapter describes another experiment that employed an objective method to further investigate this point.

The third test looked at the localisation performance of developed decoders at off-centre listening positions. The results show the off-centre optimisation component incorporated into the design tool was able to produce a decoder with improved localisation performance in off-centre positions by reducing the loudspeaker bias effect.

Finally, the results for this experiment demonstrate the design tool is able to produce good decoders. Decoders that achieve good fitness scores have been shown to perform well in reality validating the tool's performance.

Chapter 7

Binaural Evaluation of the Developed Decoders

7.1 Introduction

This chapter describes a further experiment for investigating the localisation performance of the decoders generated using the design tool. The experiment consisted of measuring head-related transfer functions (HRTFs) of a dummy head microphone for decoded sound sources panned at different angles in the horizontal plane. From these measurements each decoder was assessed on its ability to correctly produce the interaural time and level difference cues when compared with the cues from a reference real source. Previous research has shown this to be an effective method when investigating the localisation capability of an audio reproduction system (Macpherson 1989; Mac Cabe & Furlong 1994; Theiß & Hawksford 1998; Bates, Kearney, Furlong et al. 2007).

Two separate tests were conducted mirroring the listening positions and decoders evaluated in chapter 6. In the first test, measurements were made from the central listening position. In the second test, measurements were made from a number of distributed listening positions. From each position, 7 equally spaced angles between the front and rear of the ITU system were evaluated (i.e. 0°, 30°, 60°, 90°, 120°, 150°, and 180°). The aim of this work was to produce data to support the results of the listening tests.

7.2 Experimental setup

The experiment was performed in a large concert hall at the University of Huddersfield. The surround sound rig consisted of 5 Genelec 8040A loudspeakers and was set up according the ITU standard (rear speakers at $\pm 110^\circ$). A rig radius of 2m was used with the loudspeakers at a height of 1.65m positioned at least 5m from the nearest wall so early reflections would not be a factor in the measurements. The dummy head microphone used in the test was a Neumann KU 100.

In the first test the microphone was positioned at the centre point of the surround sound rig (equidistant from the loudspeakers and at the same height). In the second test the microphone was positioned at 9 different positions in the listening area (the same as in the listening test). Measurements were made using a laptop computer equipped with a RME Fireface 400 soundcard. The soundcard operated at a 48 kHz sampling rate and 16 bit resolution.

The measurement process was controlled using Adobe Audition software. The Aurora plugins (Aurora 2008) were used to generate the measurement signal - an exponential sine sweep of 10 seconds. This type of measurement signal was chosen as it maximises the signal-to-noise ratio and is robust to time variations in the system under test (Farina 2000; Farina 2007).

7.3 Data processing

The head-related impulse responses (HRIRs) were derived from the recorded sine sweeps by convolution with an inverse sine sweep signal, a process known as deconvolution. An example HRIR is shown in figure 7-1 for a real source at 0 degrees. The reverberation from the room is included in this example.

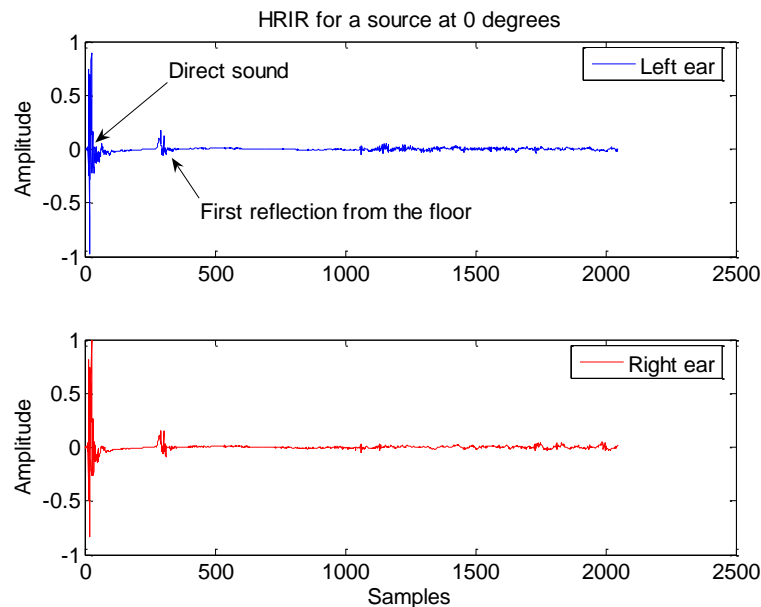


Figure 7-1: HRIR for a real source at 0 degrees

Before calculating the auditory cues the room reflections were truncated from the HRIRs. A rectangular window function was applied before the arrival of the first reflection rendering them pseudo-anechoic. After truncating the files their length was 256 samples (figure 7-2 shows the same impulse responses as figure 7-1 but with the room reflections eliminated).

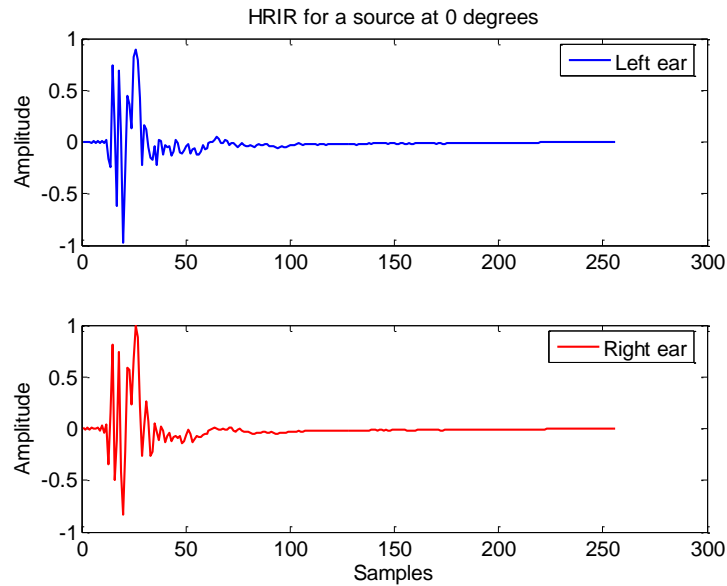


Figure 7-2: Truncated HRIR for a source at 0 degrees.

7.4 Estimation of the auditory cues using an auditory model

A computational model was developed for simulating the human auditory system and estimating the interaural cues from the measured HRIRs. The model was specifically designed to be able to give an estimation of the ITD and the ILD for HRIRs measured at the central listening position and off-centre listening positions. Its design is similar to other previously developed auditory models which have shown good agreement with subjective perception (Macpherson 1989; Theiß & Hawksford 1998; V. Pulkki & Hirvonen 2005). However, it differs by including a simple precedence effect processor at a high level of the model for suppressing later arriving wavefronts. A schematic diagram showing each processing stage of the model is presented in figure 7-3.

The first stage of the model simulates the outer ear's response to sound. This is described by the measured left ear and right ear HRTFs. Please note the middle ear has not been included because its effect is not required when estimating the two interaural cues of interest in this work (Blauert 2005).

Following this stage, the frequency analysis of the basilar membrane is simulated by filtering the left ear and right ear signals with a 36-band gammatone filterbank (Patterson et al. 1995). The centre frequencies of the filterbank follow the ERB (equivalent rectangular band-width) scale which represents the human auditory pitch scale (Brian Moore 2003).

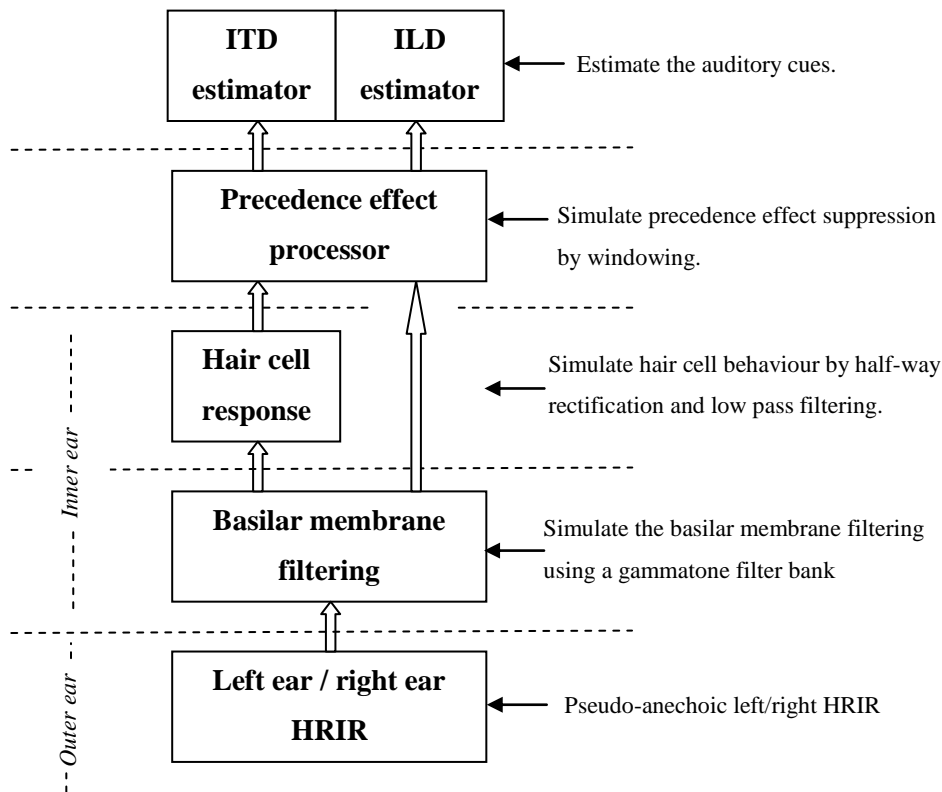


Figure 7-3: Structure of the auditory model used for estimating the interaural cues

The resulting basilar membrane signals are then transformed to neural impulses by hair cells. This is simulated at each frequency band by first compressing the signals by raising them by a power of 0.23, then subjecting them to half-way rectification and low pass filtering (filter cutoff

frequency 425 Hz). This method of hair cell modeling was proposed by Bernstein et al (Bernstein et al. 1999). Both of the inner ear signal processing stages described above were conducted using Auditory Toolbox developed by Slaney (Slaney n.d.) and Akeroyd (Akeroyd n.d.).

After the outer and inner ear stages the signals are subject to a simple precedence effect processor. A 4 ms exponential window function is centered on the first neural impulse peak from the inner ear stage. This was included to simulate localisation suppression after 1-2ms. The resulting signals were then ready for interaural time difference and interaural level difference estimation.

Research has shown that the ability of the auditory system to determine ITDs can be modeled by performing a cross-correlation on ear signals and computing the time lag at which the correlation is maximised (Blauert & Cobben 1978). In the literature this method is referred to as the Interaural Cross-Correlation (IACC) and has been used in several similar investigations (Macpherson 1989; Mac Cabe & Furlong 1994; Ville Pulkki et al. 1999; V. Pulkki & Hirvonen 2005; Bates, Kearney, Furlong et al. 2007; Pocock 1982; Jackson et al. 2008). In this work the IACC at each frequency band f and time t is defined:

$$IACC(t, f) = \max_{\tau} \left(\frac{\sum_{n=1}^N x_L(t+n)x_R(t+n+\tau)}{\sqrt{\sum_{n=1}^N x_R^2(t+n)} \sum_{n=1}^N x_R^2(t+n)} \right) \quad (7.1)$$

where τ is the lag between the two neural impulse signals (x_L and x_R) in samples (limited to the range of ± 1 ms in this work). The maximum of this function is the ITD for the corresponding frequency band. A final estimate of the ITD was computed by taking a mean value of ITD over the frequency bands between 20 – 700 Hz.

Interaural Level Differences (ILDs) were calculated by computing the ratio of energies between the left and right ear basilar membrane signals and converting to dB (see equation 7.2).

$$ILD = 10 \log_{10} \left[\frac{AP_{BasilarLeft}}{AP_{BasilarRight}} \right] \quad (7.2)$$

where the average power, AP is the energy over the duration T of the response x :

$$AP = \frac{\int_0^T x(t)^2 dt}{T} \quad (7.3)$$

The frequency dependent ILD was then averaged over all frequency bands.

7.4.1 Real source results

Finally, figure 7-4 and figure 7-5 respectively show the estimated ITD and ILD of the Neumann KU100 dummy head computed for real sound sources using the developed model for the central listening position. It can be seen that as sound moves off-axis the ITD and ILD increases. The exception is the characteristic dip at $\pm 90^\circ$ for ILD which is due to the sound waves diffracting around the head and adding in phase at the opposite ear (Blauert 2001). In figure 7-4 the ITD for a spherical head approximated using equation 2.1 was included for comparison.

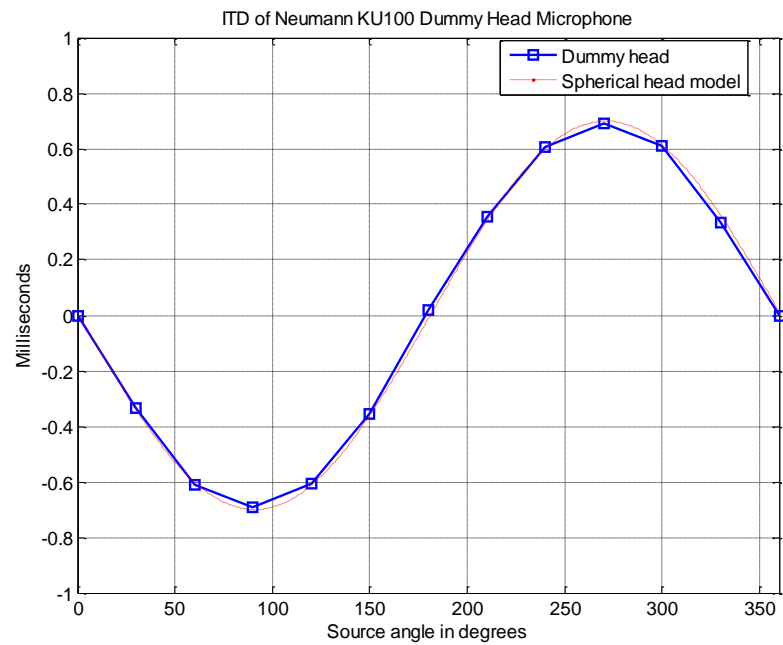


Figure 7-4: ITD of the Neumann dummy head calculated using the developed auditory model.

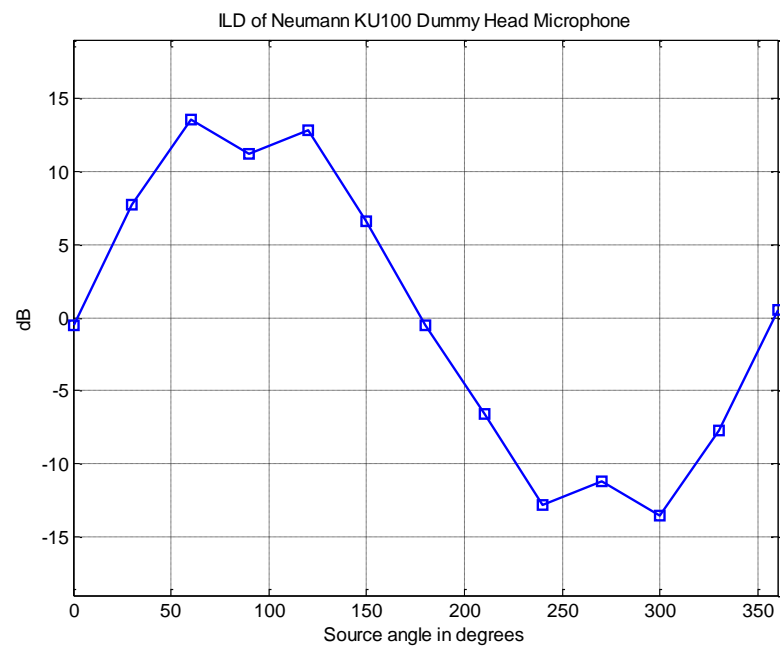


Figure 7-5: ILD of the Neumann dummy head calculated using the developed auditory model.

The trends presented in these plots are consistent with other work confirming the experimental setup, environment and the auditory model (minus the precedence effect processor). The results for the decoders will now be presented.

7.5 Test 1 - Central listening position measurements

7.5.1 Results

7.5.1.1 ITD

Figure 7-6 shows the estimated ITD for the eight decoders measured in the test. In each subplot the estimated ITD for a real source is included for reference. The ITD error (difference between the reference and the decoder) is also shown in each subplot.

Overall, Decoder 2 (first order frequency dependent) was the closest match to the reference. This was closely followed by Decoder 1 (first order frequency independent). Both decoders had good performance at the front of the system and were able to maintain good performance at the sides and at the rear. Decoder 2 in particular produced the best performance overall at the problematic angles of ± 150 . This result matches the performance predicted by the velocity vector.

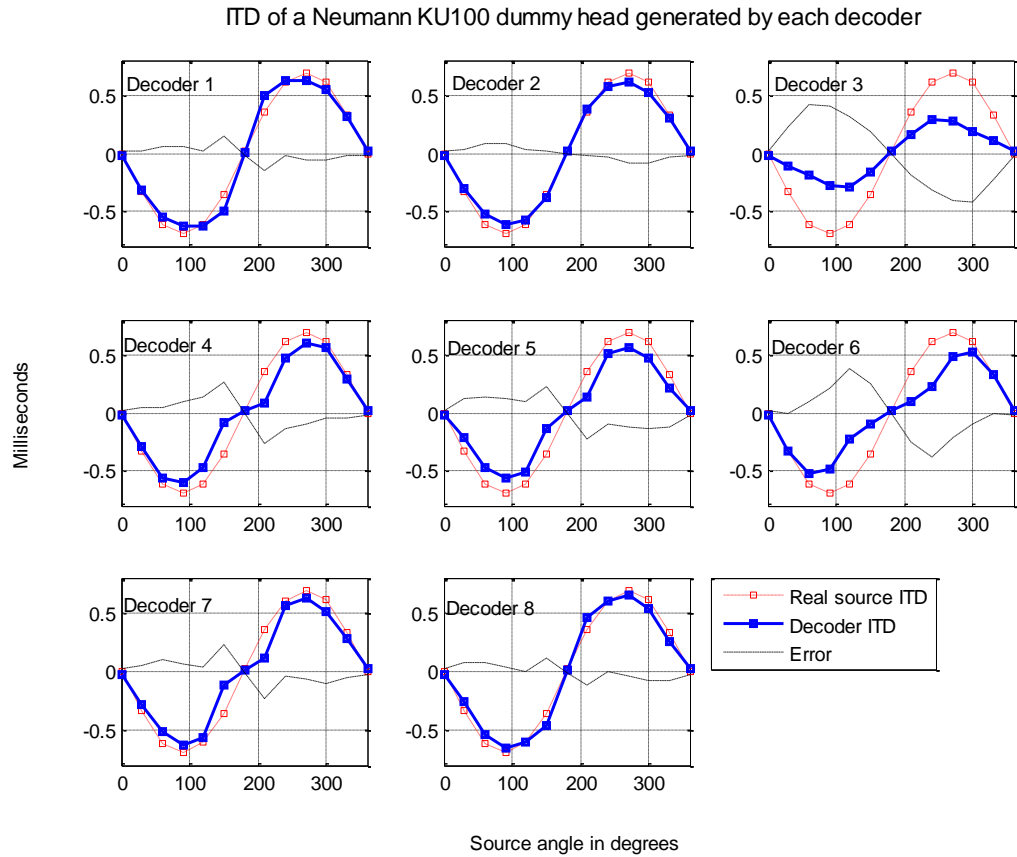


Figure 7-6: ITD for each decoder at the centre listening point (blue solid line). The ITD for a real source is included for reference (red dashed line). The error is shown as a black dash-dot line.

Decoder 3 (default Soundfield) was the worst overall (also predicted by the velocity vector). The plot for this decoder shows that ITDs at the sides are significantly lower than the reference. This suggests listeners would perceive sources as either pulling towards the front or to the rear.

ITDs were also lower than the reference at the sides for Decoders 4, 5 and 7 but to a much lesser extent. These decoders were evenly matched overall with Decoder 4 (standard fourth order) producing better performance at the front and Decoder 7 (fourth order by Wiggins) producing better performance at the rear. This result also agrees somewhat with the velocity vector predictions.

Out of all the fourth order decoders, Decoder 8 (Craven decoder) was the closest match to the performance level of Decoders 1 and 2.

The fourth order MAA optimised decoder (Decoder 6) has good frontal performance however it appears to have fallen victim to what it was aiming to achieve. Specifically, performance to the sides is degraded considerably when compared to the other fourth order decoders (note ± 120 degrees in particular). Performance was permitted to degrade at the sides but this level was not anticipated to occur prior to the tests. With regard to performance at the rear, there has been no apparent improvement. However, as pointed out earlier, it is difficult for the design tool to generate good localisation performance in the rear area of the ITU system for a frequency independent decoder given the large angular spacing between the loudspeakers.

The fourth order even performance optimised decoder (Decoder 5) has a reasonably consistent ITD error around the ITU system. However, this consistency is lost at the rear as it was in the listening test. In order to achieve even performance at the front and sides there has been an overall decrease in performance as predicted.

The most problematic area of the system for all decoders (apart from Decoder 2 and Decoder 3) is at the rear. The plots suggest that for most decoders phantom images generated at ± 150 degrees will pull away from the intended direction. Specifically, phantom images for Decoder 1 (first order frequency independent) and Decoder 8 (fourth order Craven decoder) will pull towards the side of the listener (higher ITD), whereas phantom images for the other fourth order decoders and Decoder 3 will pull towards the direct rear of the listener (lower ITD).

Interestingly, there is some evidence to suggest this observation correlates with the results obtained in the listening test. To illustrate this point, figure plots the mean subject response in the listening test for the low frequency noise source at the angles of 0° , 30° , 60° , 90° , 120° , 150° and 180° (without front-back reversals). Note the correlation between the plots in figure and figure

particularly at 150 degrees. For example, at 150 degrees Decoder 8 has a higher ITD and the mean subject response draws to the side.

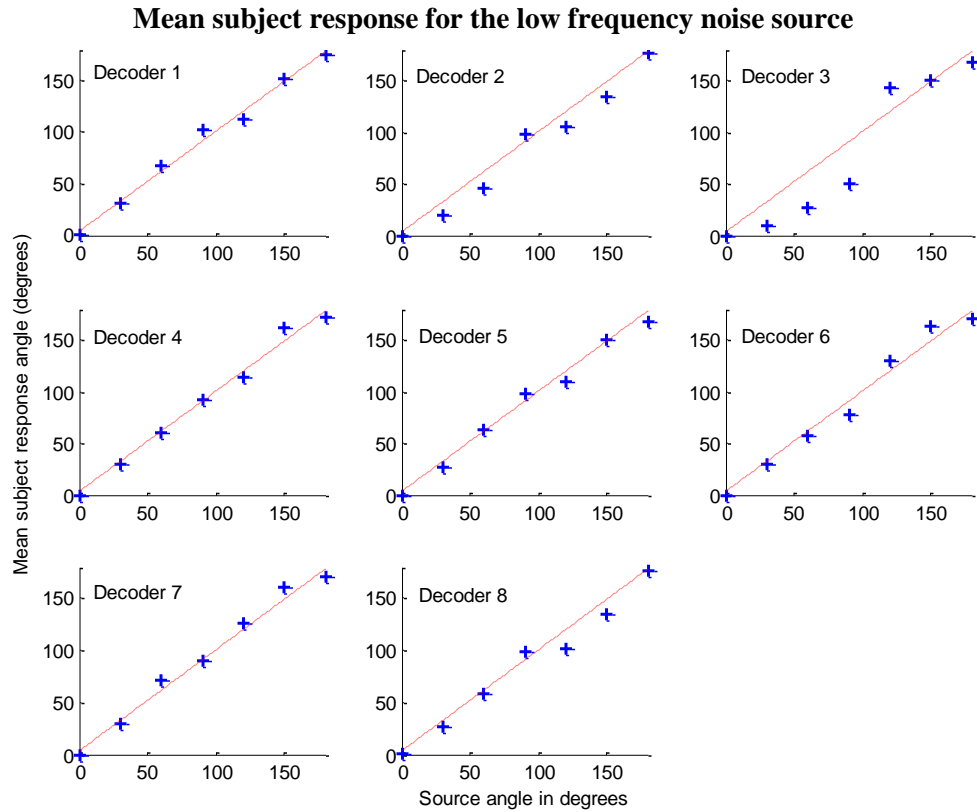


Figure 7-7: Mean subject response in the listening test for the low frequency noise source. The red dashed line indicates the ideal response.

7.5.1.2 ILD

Figure 7-8 plots the estimated ILDs for each decoder. In this figure the results suggest that the angles of ± 60 degrees and ± 120 degrees were the most problematic in terms of generating accurate ILDs. At both of these angles most decoders were unable to produce level differences significant enough to match the real source data implying that a centrally seated listener would perceive sources as pulling towards the front or the rear. This problem is most noticeable for the first order decoders (Decoders 1, 2 and 3) and the MAA optimised decoder (Decoder 6).

Overall Decoder 8 (fourth order Craven decoder) was the closest match to the reference. This was closely followed by Decoder 5 (even error optimised) and Decoder 7 (fourth order by Wiggins). When compared to all other decoders, Decoder 8 excels in performance at ± 150 degrees. This is an interesting result considering the poor performance predicted by the energy vector at the rear of the system. Specifically, the energy vector angle pulls away from the intended angle towards the rear loudspeakers (see the performance plot in Appendix B).

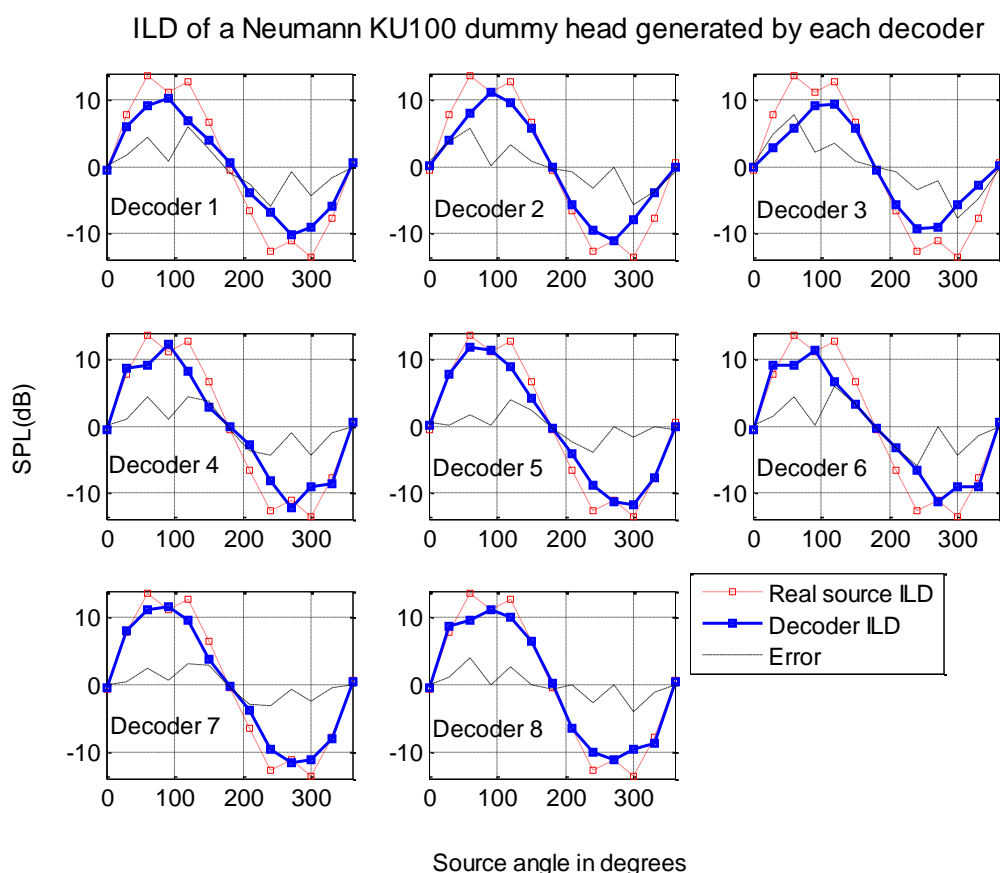


Figure 7-8: ILD for each decoder measured from the centre listening point (blue solid line). The ILD for a real source is included for reference (red dashed line).

Decoder 1 (first order frequency independent) and Decoder 2 (first order frequency dependent) were quite closely matched overall. Decoder 1 performs better in the front half of the sound stage

whereas Decoder 2 performs better in the rear half of the sound stage. Decoder 3 (default Soundfield) is comparable to Decoder 2 at the rear but worse than both Decoders 1 and 2 at the front.

Decoder 4 (fourth order range-removal decoder) is fairly close to the level of Decoder 5 and Decoder 7 except for the larger error at the side angles of ± 60 degrees and the rear-side error at ± 150 .

The frontal performance for Decoder 6 (fourth order MAA) is comparable with Decoder 4 and Decoder 8. However, it appears that performance to the rear and sides has suffered (note the angles of 120 degrees and 150 degrees in particular).

When considering the relative differences between the decoders in this plot there is strong agreement with the energy vector predictions. The first order decoders were predicted to perform worse than the fourth order decoders and this was the case.

7.5.2 Discussion

In general the results of this test concur with the performance predictions made by the velocity vector and energy vector. For example, Decoders 1 and 2 both had ideal velocity vector responses resulting in ITDs that closely matched the ITDs of a real source at the angles evaluated. In addition, the predicted good performance at the front and poor performance at the sides can be seen for the MAA optimised decoder (Decoder 6).

One notable exception was the good ILD performance of the fourth order Craven decoder at the rear. Although the energy vector magnitude for this decoder was comparable to the other decoders under test, the energy vector angle was particularly poor at the rear. This suggests that the vectors cannot be completely relied upon when estimating performance between a pair of widely spaced loudspeakers. The good performance for this decoder was also noted by Wiggins during a similar analysis of ILD performance (Wiggins 2007).

The results from this analysis show that, on the whole, the optimisation components added to the design tool were successful. Good frontal performance was predicted and achieved for the MAA optimised decoder in both the ITD and ILD plots (although this good level of performance has been at the expense of performance to the sides and rear to a greater level than anticipated prior to the test).

Even performance was achieved for Decoder 5 (especially around the front half of the system as predicted). However, striving for even performance has resulted in a reduced level of ITD performance around the sound stage when compared to the other fourth order decoders. To improve even performance and overall performance for both the velocity vector and the energy vector a dual-band decoder could be implemented.

Whilst these results appear less conclusive than the listening tests in terms of matching ranked order of decoder performance to predicted performance, it is good they show that the decoders produced by the design tool perform at a similar level to the best published decoders, especially bearing in mind that Wiggins' decoder was derived and selected on the basis of its good performance under HRTF analysis and Craven's decoder was carefully hand-crafted.

7.6 Test 2 - Off-centre listening position measurements

7.6.1 Results

7.6.1.1 ITD

Table 7.1 presents the unsigned mean ITD error by angle for each decoder for each measurement position (the error in this case is defined as the difference between the ITD of the real source and the ITD of the decoded source). The mean (μ) and standard deviation (σ) of the error is shown for each decoder in each position. To guide the reader, better performance is indicated by shaded cells.

From this data it can be seen the off-centre optimised decoder was able to produce the best performance at 5 out of 9 listening positions (front-left, left, right, back and front-right) and equal best at the back-right position. The decoder by Poletti performed the best in the remaining 2 off-centre positions but was very close overall to the off-centre optimised decoder in most positions.

The main area of the system in which these decoders performed best was at the front. Figure 7-9 to figure 7-17 show this by plotting the estimated ITD at each of the listening positions. Note that both decoders (Decoders 3 and 5) produced ITDs at the front of the sound stage that are much closer to the ideal response line at most positions. Exceptions to this are measurement made from the front-left position where both decoders exhibit a large error at 90 degrees (see data in table 7.1 and figure 7-11). In addition, both decoders exhibit large errors as a source is at 60 degrees from the back-left position. This shows that even though both decoders come out the best in this test there are negative points to their performance.

At off-centre listening positions the first order decoder (Decoder 1) performed worse overall as expected. For example, the almost constant ITD for this decoder at the back-right position (figure 7-15) indicates that sources will be localised in the direction of the nearest loudspeaker. This is in agreement with the results presented for the listening test. In addition, at the left side listening positions the ITD is almost opposite the reference ITD at some angles.

The standard fourth order decoder (Decoder 2) and the fourth order decoder by Wiggins (Decoder 4) were similar at most listening positions. Performance-wise, both decoders are better than the first order decoder on average at all positions apart from the centre and front-right. However, they never exceed the performance of the two off-centre optimised decoders (Decoders 3 and 5) at any position.

Mean unsigned ITD error by angle (ms)										
Position	Decoder	0	30	60	90	120	150	180	μ	σ
Centre	First order	0.02	0.02	0.06	0.06	0.02	0.15	0.01	0.05	0.05
	Standard fourth order	0.02	0.04	0.04	0.09	0.13	0.27	0.01	0.09	0.09
	Fourth order off-centre	0.02	0.03	0.08	0.06	0.04	0.25	0.01	0.07	0.08
	Fourth order Wiggins	0.02	0.06	0.10	0.07	0.04	0.23	0.01	0.08	0.08
	Fourth order Poletti	0.02	0.02	0.08	0.07	0.05	0.20	0.01	0.07	0.07
Front	First order	0.03	0.43	0.26	0.09	0.88	0.78	0.02	0.35	0.36
	Standard fourth order	0.01	0.28	0.14	0.08	0.12	0.28	0.01	0.13	0.11
	Fourth order off-centre	0.02	0.17	0.14	0.09	0.05	0.14	0.02	0.09	0.06
	Fourth order Wiggins	0.02	0.42	0.16	0.10	0.13	0.16	0.01	0.14	0.14
	Fourth order Poletti	0.02	0.02	0.15	0.17	0.05	0.05	0.02	0.07	0.06
Front-Left	First order	0.50	0.90	1.25	0.76	0.73	0.55	0.18	0.70	0.34
	Standard fourth order	0.69	0.07	0.46	0.42	0.62	0.78	0.60	0.52	0.23
	Fourth order off-centre	0.05	0.37	0.05	1.05	0.52	0.31	0.39	0.39	0.34
	Fourth order Wiggins	0.56	0.80	0.70	0.41	0.74	0.74	0.59	0.65	0.14
	Fourth order Poletti	0.14	0.56	0.27	0.82	0.61	0.32	0.09	0.40	0.27
Left	First order	0.48	0.96	1.04	0.81	0.69	0.89	0.68	0.79	0.19
	Standard fourth order	0.57	0.10	0.20	0.31	0.83	1.00	0.71	0.53	0.34
	Fourth order off-centre	0.04	0.49	0.01	0.21	0.03	0.58	0.67	0.29	0.29
	Fourth order Wiggins	0.49	1.04	0.07	0.36	0.77	0.99	0.72	0.63	0.35
	Fourth order Poletti	0.02	0.23	0.01	0.38	0.14	0.89	0.69	0.34	0.34
Back-Left	First order	0.61	1.15	0.69	0.34	0.97	1.03	0.54	0.76	0.30
	Standard fourth order	0.28	0.70	0.71	0.08	0.97	0.82	0.75	0.61	0.32
	Fourth order off-centre	0.00	0.61	0.96	0.24	0.00	0.71	0.75	0.47	0.38
	Fourth order Wiggins	0.35	0.53	0.35	0.66	0.70	0.82	0.54	0.56	0.18
	Fourth order Poletti	0.02	0.08	1.07	0.10	0.30	0.87	0.75	0.45	0.43
Back	First order	0.01	0.49	0.39	0.03	0.16	0.34	0.04	0.21	0.20
	Standard fourth order	0.00	0.10	0.15	0.04	0.21	0.39	0.04	0.13	0.14
	Fourth order off-centre	0.01	0.05	0.02	0.03	0.01	0.10	0.04	0.04	0.03
	Fourth order Wiggins	0.00	0.04	0.00	0.16	0.12	0.37	0.04	0.10	0.13
	Fourth order Poletti	0.00	0.01	0.24	0.04	0.00	0.03	0.04	0.05	0.08
Back-Right	First order	0.50	0.50	0.24	0.02	0.04	0.40	0.66	0.34	0.25
	Standard fourth order	0.19	0.09	0.24	0.02	0.03	0.40	0.45	0.20	0.17
	Fourth order off-centre	0.09	0.02	0.24	0.02	0.04	0.40	0.46	0.18	0.19
	Fourth order Wiggins	0.50	0.05	0.24	0.02	0.04	0.40	0.45	0.24	0.21
	Fourth order Poletti	0.02	0.01	0.09	0.02	0.04	0.40	0.67	0.18	0.26
Right	First order	0.49	0.19	0.05	0.04	0.50	0.40	0.76	0.35	0.27
	Standard fourth order	0.54	0.07	0.11	0.03	0.06	0.43	0.77	0.29	0.29
	Fourth order off-centre	0.13	0.03	0.11	0.07	0.10	0.46	0.77	0.24	0.27
	Fourth order Wiggins	0.48	0.19	0.06	0.24	0.11	0.46	0.78	0.33	0.26
	Fourth order Poletti	0.05	0.03	0.42	0.06	0.08	0.44	0.76	0.26	0.28
Front-Right	First order	0.53	0.00	0.39	0.42	0.20	0.60	0.37	0.36	0.20
	Standard fourth order	0.49	0.00	0.41	0.52	0.17	0.41	0.69	0.39	0.23
	Fourth order off-centre	0.07	0.01	0.42	0.03	0.09	0.36	0.65	0.23	0.25
	Fourth order Wiggins	0.51	0.00	0.40	0.54	0.15	0.43	0.67	0.39	0.23
	Fourth order Poletti	0.05	0.01	0.44	0.04	0.07	0.33	0.73	0.24	0.27

Table 7.1: Mean unsigned ITD error by angle (ms) for all decoders and all measurement positions. All decoders exhibit a higher error at the rear of the system.

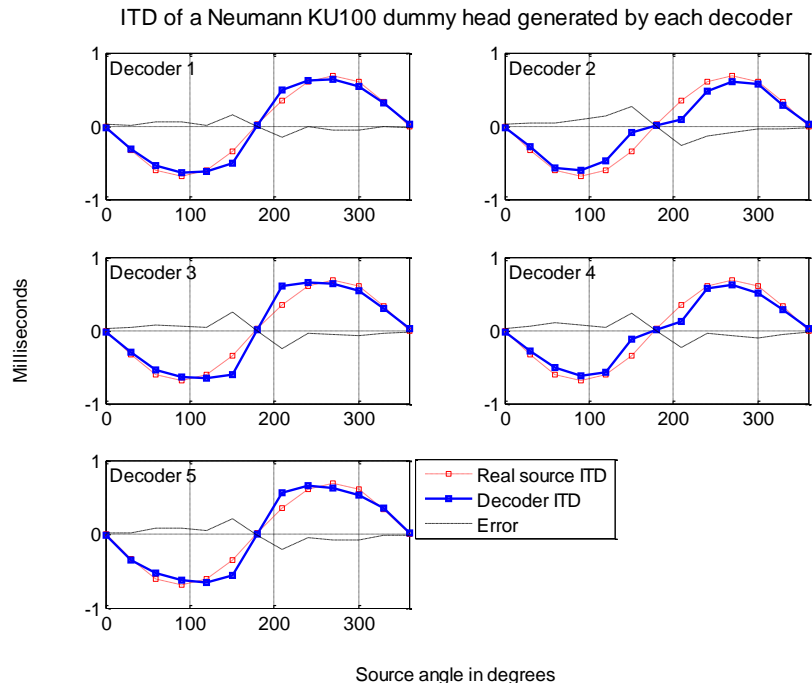


Figure 7-9: ITD by angle for off-centre tested decoders at the centre position

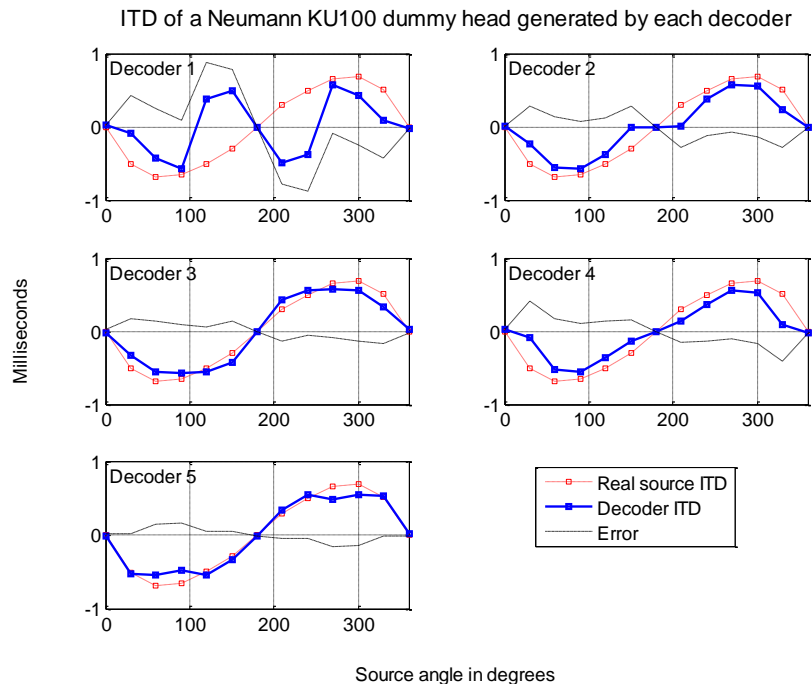


Figure 7-10: ITD by angle for off-centre tested decoders at the front position

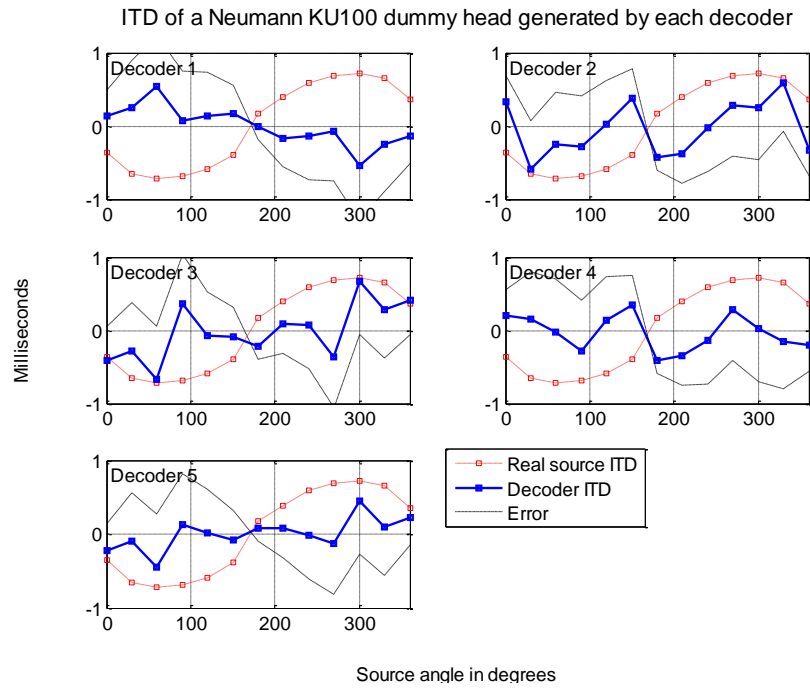


Figure 7-11: ITD by angle for off-centre tested decoders at the front-left position

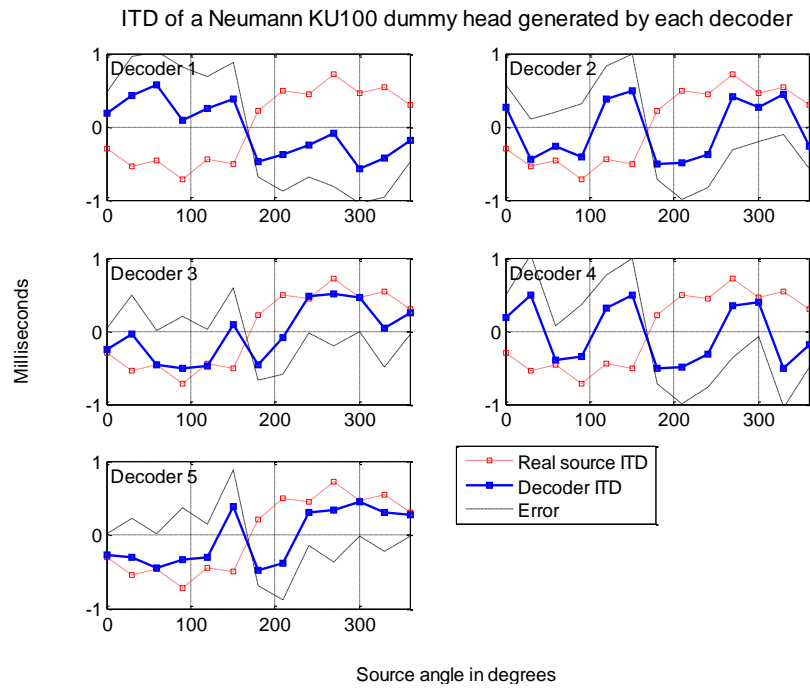


Figure 7-12: ITD by angle for off-centre tested decoders at the left position

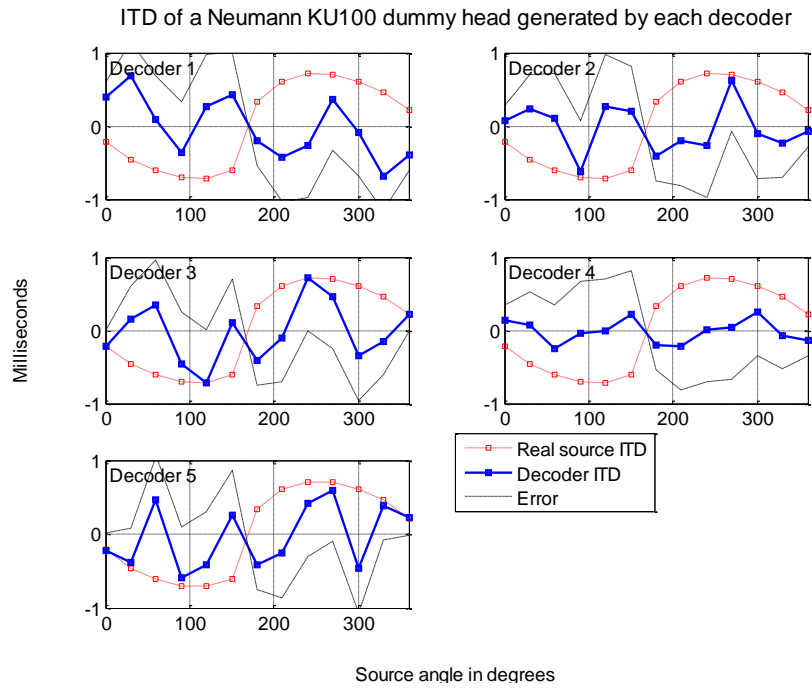


Figure 7-13: ITD by angle for off-centre tested decoders at the back-left position

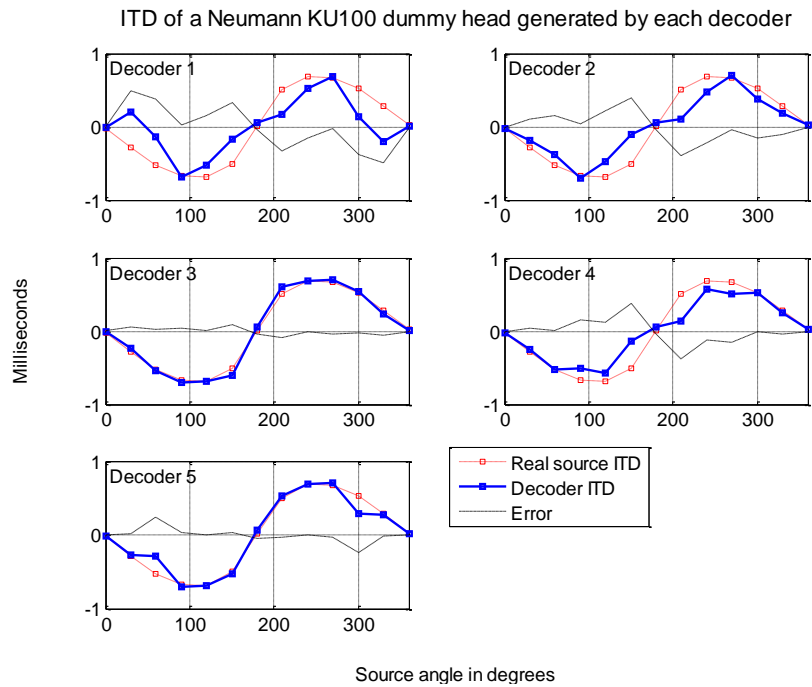


Figure 7-14: ITD by angle for off-centre tested decoders at the back position

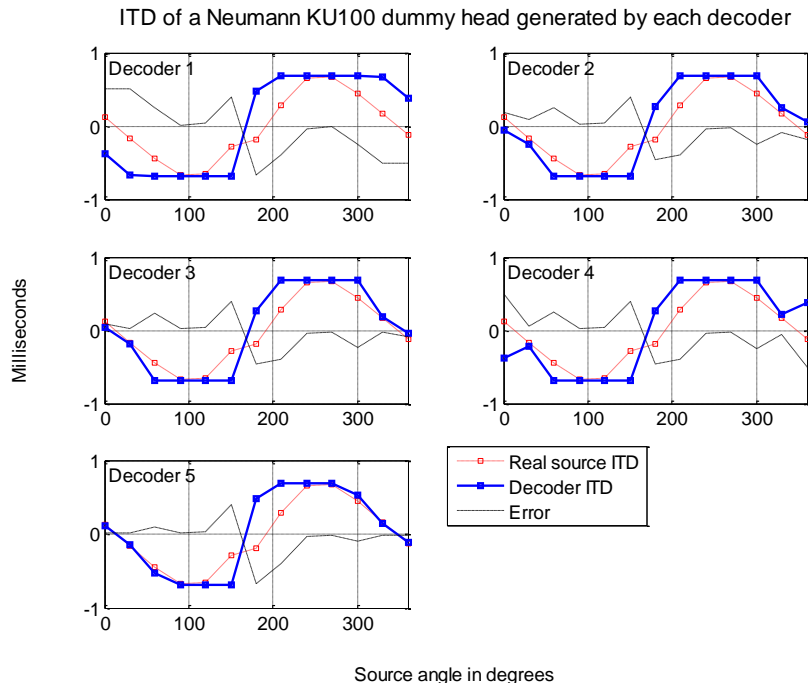


Figure 7-15: ITD by angle for off-centre tested decoders at the back-right position

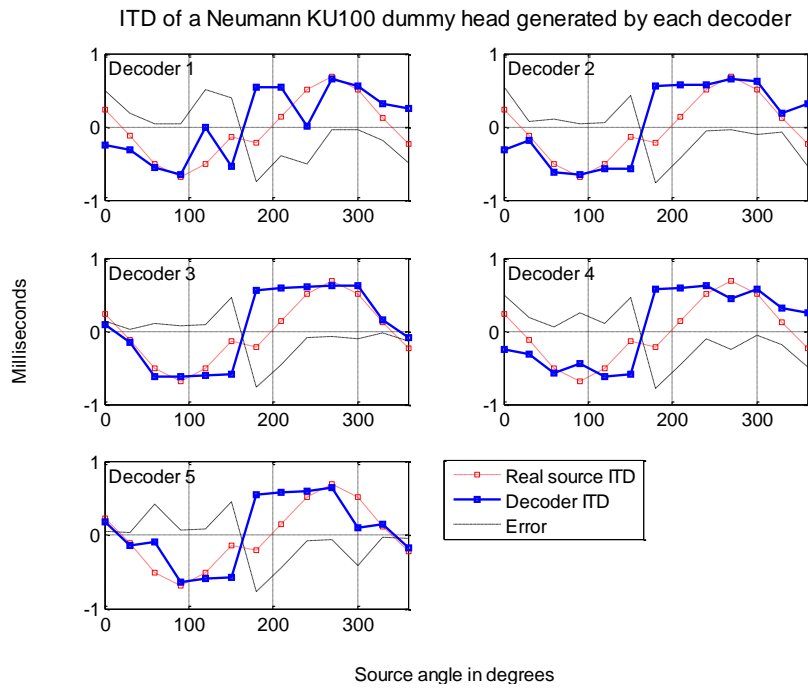


Figure 7-16: ITD by angle for off-centre tested decoders at the right position

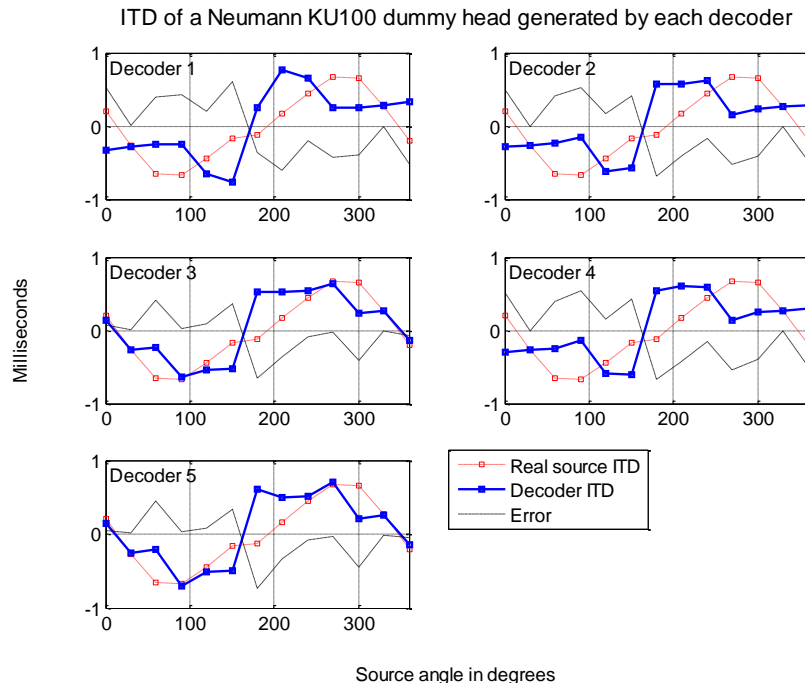


Figure 7-17: ITD by angle for off-centre tested decoders at the front-right position

Interestingly, the performance of the off-centre optimised decoder at the rear listening position is deemed by the auditory model as almost ideal. This agrees to some extent with the good performance seen for this decoder in the listening tests.

The model predicts the worst listening positions are on the opposite of the sound stage as the sound source (front-left, left and back-left) as was the case in the listening tests.

7.6.1.2 ILD

Table 7.2 presents the data for the ILD. The data shows that the off-centre optimised decoder gave the best performance at the left, back-left and front-right listening positions, the decoder by Poletti gave the best performance at front-left and back listening positions and, interestingly, the non off-centre optimised decoder by Wiggins performed the best in the remaining positions (centre, front, back-right, right). In order to provide further information figure 7-18 to figure 7-26 show the ILD by angle for all decoders at each listening position.

Mean unsigned ILD error by angle (dB)										
Position	Decoder	0	30	60	90	120	150	180	μ	σ
Centre	First order	0.01	1.66	4.32	0.88	5.95	2.59	1.09	2.36	2.10
	Standard fourth order	0.07	1.04	4.31	1.02	4.46	3.82	0.46	2.17	1.94
	Fourth order off-centre	0.19	1.58	3.88	1.54	0.11	3.04	0.65	1.57	1.44
	Fourth order Wiggins	0.03	0.38	2.45	0.57	3.13	2.82	0.35	1.39	1.34
	Fourth order Poletti	0.30	0.80	6.00	1.46	0.45	0.85	0.70	1.51	2.01
Front	First order	0.34	7.36	2.39	5.89	6.76	3.88	0.24	3.84	2.96
	Standard fourth order	0.35	4.00	0.94	2.06	3.23	2.13	0.74	1.92	1.35
	Fourth order off-centre	0.34	5.51	0.75	1.90	1.38	4.67	0.56	2.16	2.09
	Fourth order Wiggins	0.99	4.18	0.42	2.75	1.54	1.57	0.83	1.75	1.30
	Fourth order Poletti	0.35	2.78	2.06	3.03	1.31	3.17	0.54	1.89	1.18
Front-Left	First order	8.62	11.53	2.13	9.00	13.16	11.43	6.74	8.94	3.70
	Standard fourth order	4.50	2.27	0.76	0.35	14.70	13.67	9.43	6.52	6.05
	Fourth order off-centre	0.15	5.10	1.33	1.81	1.44	8.45	8.45	3.82	3.51
	Fourth order Wiggins	7.18	5.64	1.64	1.46	12.62	13.88	9.81	7.46	4.94
	Fourth order Poletti	0.24	3.40	1.24	5.89	5.72	1.34	8.18	3.72	2.96
Left	First order	5.55	8.59	7.50	2.64	9.89	10.09	6.04	7.18	2.66
	Standard fourth order	3.20	1.49	4.44	0.16	9.86	11.33	6.49	5.28	4.18
	Fourth order off-centre	0.33	2.39	3.25	0.61	0.11	1.48	6.19	2.05	2.16
	Fourth order Wiggins	4.73	3.21	3.03	0.15	6.73	10.75	6.49	5.01	3.39
	Fourth order Poletti	0.28	1.08	3.46	0.54	0.16	3.95	6.36	2.26	2.37
Back-Left	First order	5.70	12.73	10.75	6.19	9.85	13.95	10.64	9.97	3.08
	Standard fourth order	1.27	4.88	7.09	3.58	10.46	15.48	10.76	7.65	4.90
	Fourth order off-centre	0.41	1.99	3.52	2.88	0.74	6.06	10.71	3.76	3.61
	Fourth order Wiggins	4.12	6.36	3.21	4.41	7.61	14.90	10.72	7.33	4.20
	Fourth order Poletti	0.54	0.68	1.25	3.93	1.38	8.23	10.74	3.82	4.09
Back	First order	1.10	4.64	6.01	5.51	7.50	7.29	0.39	4.63	2.84
	Standard fourth order	0.84	3.06	4.17	4.15	7.11	8.60	0.35	4.04	3.03
	Fourth order off-centre	0.57	3.76	3.60	3.55	2.76	2.37	0.35	2.42	1.43
	Fourth order Wiggins	1.03	2.88	3.73	4.51	5.64	7.82	0.33	3.71	2.60
	Fourth order Poletti	0.55	1.70	3.36	3.92	2.95	3.64	0.34	2.35	1.48
Back-Right	First order	2.43	1.54	2.46	4.71	5.56	0.72	6.96	3.48	2.29
	Standard fourth order	1.70	0.79	2.46	3.53	4.68	0.21	7.35	2.96	2.47
	Fourth order off-centre	0.09	1.75	2.53	2.12	0.83	5.59	7.20	2.87	2.58
	Fourth order Wiggins	2.23	0.09	1.79	3.94	3.34	0.72	7.42	2.79	2.45
	Fourth order Poletti	0.10	0.65	4.84	2.69	1.12	4.17	7.16	2.96	2.57
Right	First order	5.42	1.49	5.57	2.25	2.73	3.84	7.17	4.07	2.06
	Standard fourth order	3.19	0.56	6.32	1.03	1.40	3.60	7.58	3.38	2.70
	Fourth order off-centre	0.49	1.20	6.47	0.57	2.63	9.37	7.49	4.03	3.67
	Fourth order Wiggins	4.69	0.36	5.20	1.45	0.07	4.12	7.61	3.36	2.81
	Fourth order Poletti	0.44	0.05	8.43	0.08	2.19	7.84	7.46	3.78	3.94
Front-Right	First order	7.39	3.20	8.91	6.84	3.31	0.47	3.61	4.82	2.96
	Standard fourth order	3.46	2.01	8.43	5.88	1.24	1.32	4.22	3.79	2.64
	Fourth order off-centre	0.09	2.60	8.37	4.33	2.22	6.46	0.00	3.44	3.15
	Fourth order Wiggins	5.85	2.13	8.14	6.49	0.43	1.85	4.12	4.14	2.81
	Fourth order Poletti	0.02	1.10	9.01	4.71	2.02	4.93	4.26	3.72	3.00

Table 7.2: Mean unsigned ILD error by angle (dB) for all decoders and measurement positions.

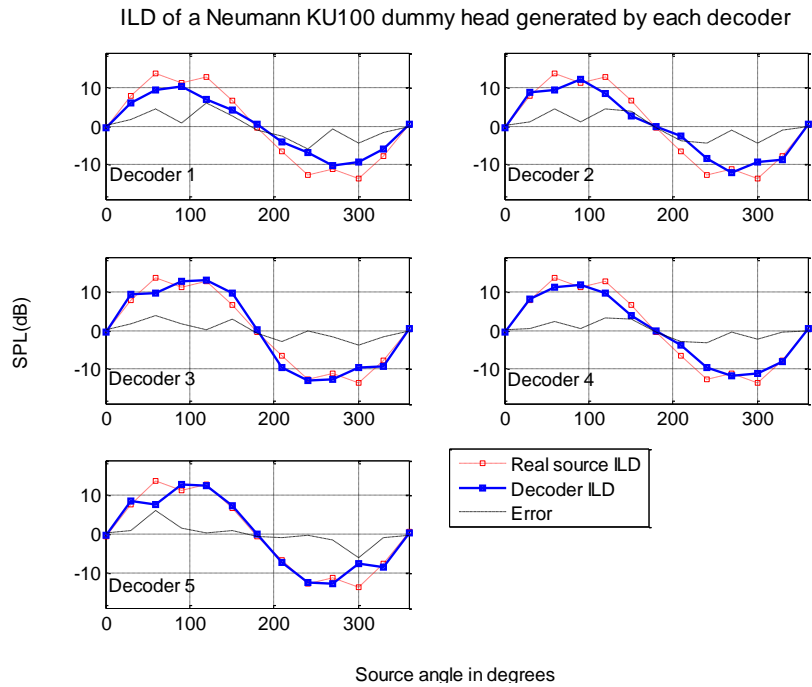


Figure 7-18: ILD by angle for off-centre tested decoders at the centre position

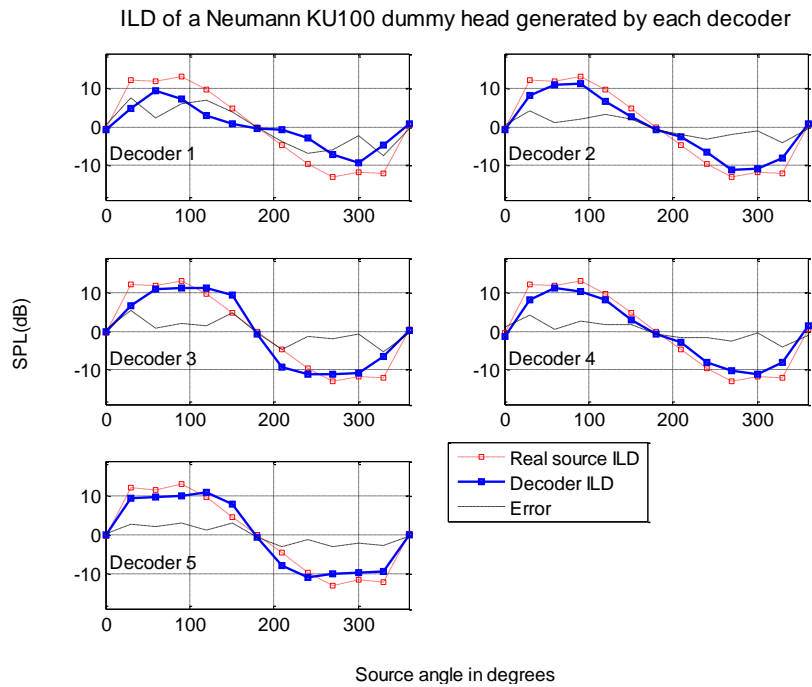


Figure 7-19: ILD by angle for off-centre tested decoders at the front position

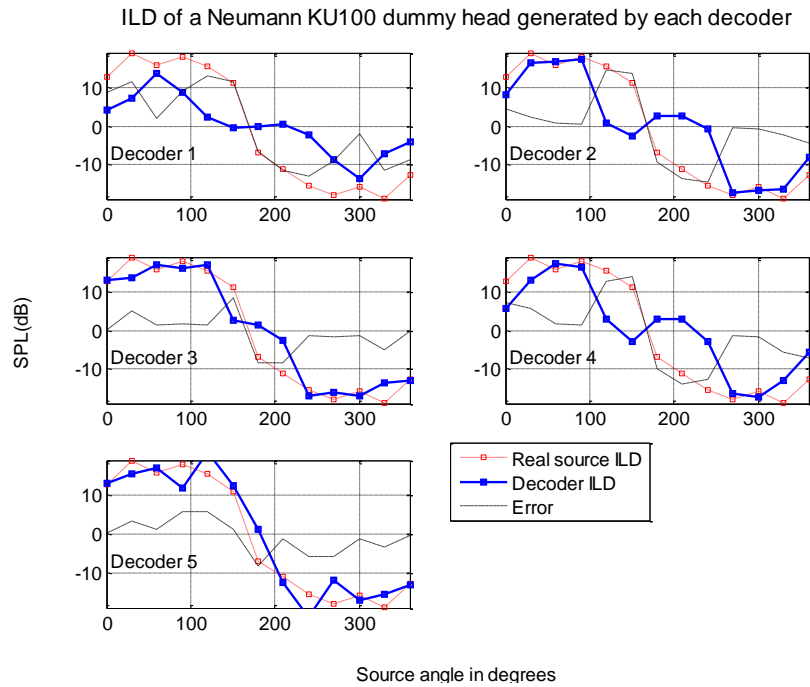


Figure 7-20: ILD by angle for off-centre tested decoders at the front-left position

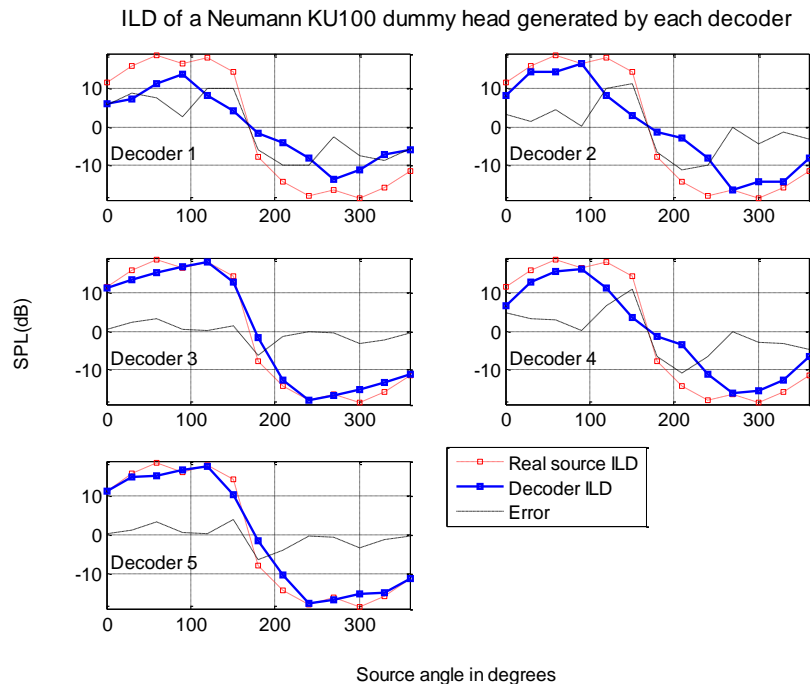


Figure 7-21: ILD by angle for off-centre tested decoders at the left position

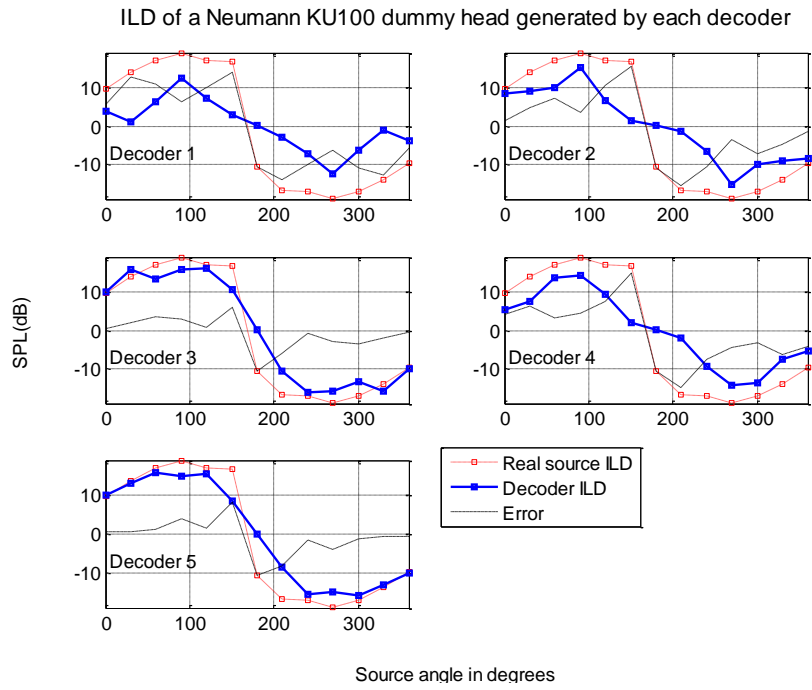


Figure 7-22: ILD by angle for off-centre tested decoders at the back-left position

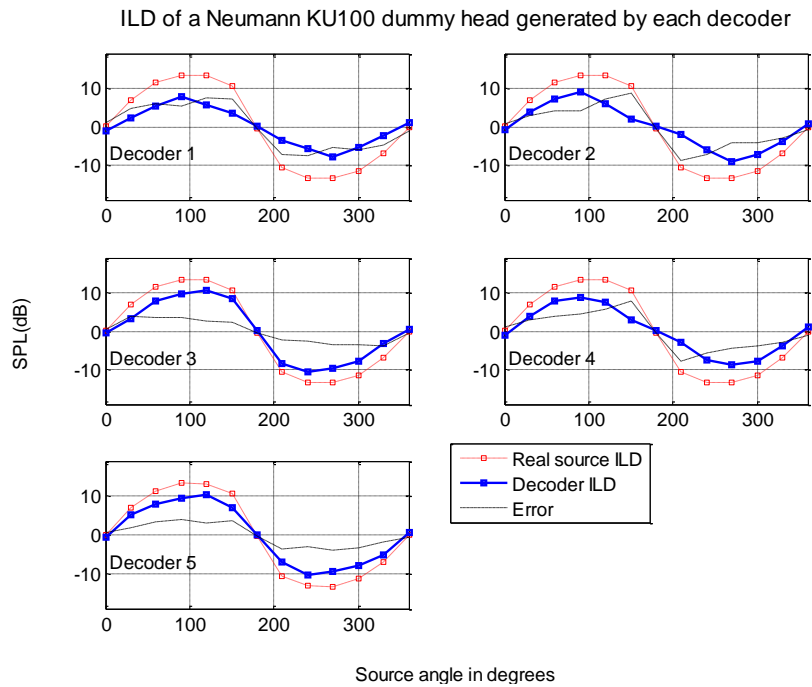


Figure 7-23: ILD by angle for off-centre tested decoders at the back position

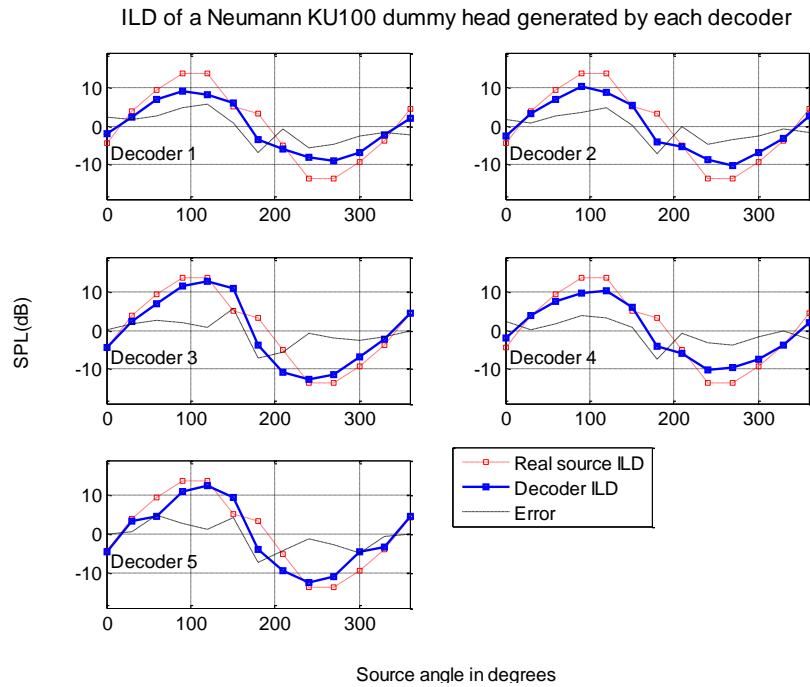


Figure 7-24: ILD by angle for off-centre tested decoders at the back-right position

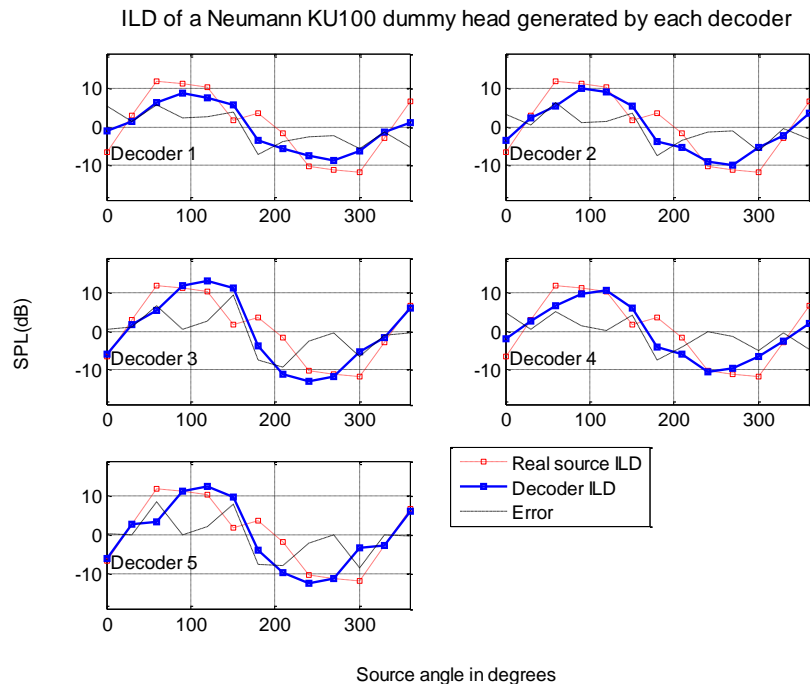


Figure 7-25: ILD by angle for off-centre tested decoders at the right position

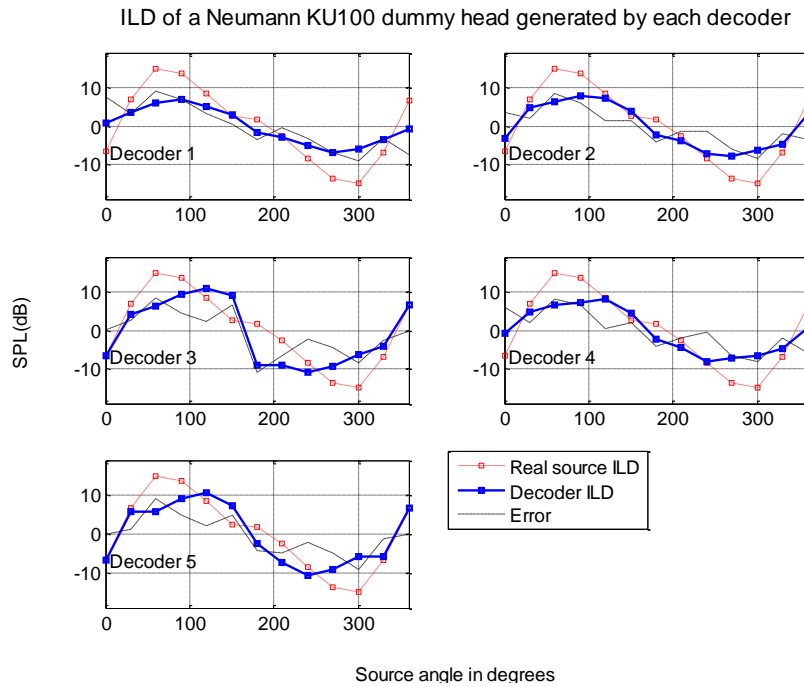


Figure 7-26: ILD by angle for off-centre tested decoders at the front-right position

At the central listening position the decoder by Wiggins (Decoder 4) is closest fit to the reference (see figure 7-18). This level of performance and consistency around the sound stage is also maintained for the front listening position (see figure 7-19).

When examining the plots for the most problematic listening positions (front-left, left and back-left) the off-centre optimised decoder and the decoder by Poletti perform significantly better especially when positioning sources to the front.

In all positions where the ITU array is left-right symmetrical about the listener, the difference in performance between the decoders is not as significant and also the error is generally lower.

7.6.2 Overall decoder performance

Table 7.3 and table 7.4 present the mean overall ITD and ILD error for each measurement position for each decoder. A total mean error is included as well as the standard deviation of the error in the right hand columns of the tables. The results show that the off-centre optimised

decoder and the decoder by Poletti performed the best and were very evenly matched. Both decoders were able to outperform the other decoders in terms of producing the lowest overall error and were the most consistent around the sound stage (indicated by the standard deviation).

As previously highlighted, the non off-centre optimised decoder by Wiggins performed well in terms of delivering good ILD performance at 4 listening positions (centre, front, back-right and right).

In both tables the mean error at each listening position (average of all decoders) shows that the best listening positions as the centre (as expected) and the worst was the back-left position on the opposite side of the sound stage as the source (agreeing with the listening test results).

	Overall mean unsigned ITD error (ms)										
Decoder	Centre	Front	Front-Left	Left	Back-Left	Back	Back-Right	Right	Front-Right	μ	σ
First order	0.05	0.35	0.70	0.79	0.76	0.21	0.34	0.35	0.36	0.43	0.26
Standard fourth order	0.09	0.13	0.52	0.53	0.61	0.13	0.20	0.29	0.39	0.32	0.20
Fourth order off-centre	0.07	0.09	0.39	0.29	0.47	0.04	0.18	0.24	0.23	0.22	0.15
Fourth order Wiggins	0.08	0.14	0.65	0.63	0.56	0.10	0.24	0.33	0.39	0.35	0.23
Fourth order Poletti	0.07	0.07	0.40	0.34	0.45	0.05	0.18	0.26	0.24	0.23	0.15
μ	0.07	0.16	0.53	0.52	0.57	0.11	0.23	0.29	0.32		

Table 7.3: Mean ITD error for each decoder at each listening position. The mean and standard deviation are included in the right hand columns

	Overall mean unsigned ILD error (dB)										
Decoder	Centre	Front	Front-Left	Left	Back-Left	Back	Back-Right	Right	Front-Right	μ	σ
First order	2.36	3.84	8.94	7.18	9.97	4.63	3.48	4.07	4.82	5.48	2.61
Standard fourth order	2.17	1.92	6.52	5.28	7.65	4.04	2.96	3.38	3.79	4.19	1.94
Fourth order off-centre	1.57	2.16	3.82	2.05	3.76	2.42	2.87	4.03	3.44	2.90	0.90
Fourth order Wiggins	1.39	1.75	7.46	5.01	7.33	3.71	2.79	3.36	4.14	4.11	2.18
Fourth order Poletti	1.51	1.89	3.72	2.26	3.82	2.35	2.96	3.78	3.72	2.89	0.91
μ	1.80	2.31	6.09	4.36	6.51	3.43	3.01	3.72	3.98		

Table 7.4: Mean ILD error for each decoder at each listening position. The mean and standard deviation are included in the right hand column.

7.6.3 Discussion

Binaural measurements were made from off-centre listening positions in order to estimate and compare the auditory cues for a real source and decoded sources. The results presented are in general agreement with the listening test results in that the fourth order off-centre optimised decoder and the decoder by Poletti were able to produce the best performance. Specifically, the overall decoder ranking, derived from the total means in table 7.3 and table 7.4 generally agrees with the rankings for the low and mid/high frequency noise listening tests.

The auditory model incorporated into this work for predicting localisation in off-centre listening appears to work well. The application of this model has never been used in this area of research before and warrants further investigation.

7.7 Summary

This chapter further investigated the localisation performance of the developed decoders by means of a binaural analysis experiment. The listening positions measured mirror those evaluated in the listening test and the data provided supports the listening test results which demonstrate that the components added to the tool are generally successful in meeting their overall objectives.

Chapter 8

Further optimisation of existing Ambisonic decoders

8.1 Introduction

After the listening tests two additional experiments were carried out using decoders derived by other authors as starting points for the search. The aim was to see if existing decoders derived using different techniques could be improved using the design tool. In the first application the fourth order decoder by Craven was used as a starting point. In the second application the fourth order decoder by Poletti was used as a starting point. These decoders were selected on the basis of their comparatively good performance in the tests described in chapter 6 and chapter 7. The following sections examine the performance of the best solution derived in each application.

8.2 Further optimisation of the Craven decoder

Craven aimed to meet the following design criteria when deriving his fourth order decoder (Craven 2003):

- Reproduced energy should be substantially independent of panning angle.
- The velocity and energy vector directions θ_v θ_E should be closely matched.
- The angles θ_v θ_E should be reasonably close to the panning angle θ .
- Velocity vector length r_v should be close to unity.
- Energy vector length r_E should be as large as possible.

Information about exactly how the above criteria were formulated in a fitness function was not detailed in Craven's work. However, Craven states that a conjugate-gradient search algorithm was employed for locating his decoder in the search space. When using the design tool only the range-removal component was used. Importance weightings were chosen to broadly reflect Craven's design criteria (see table 8.1).

E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
0.0	10.0	1.0	1.0	1.0	1.0	1.5
$E_{LFAngEv}$	$E_{HFAngEv}$	$E_{LFMagEv}$	$E_{HFMagEv}$			
0.1	0.1	0.1	0.1			

Table 8.1: Importance weights used when deriving a solution from Craven's decoder

Figure 8-1 shows the performance of the best solution produced after 10 runs of the design tool when using Craven's coefficients as a start point. For comparison, the performance of the original Craven decoder is shown in figure 8-2.

At first glance the performance of both decoders may seem alike. Indeed, the energy vector magnitudes and angles are very similar for both decoders. However, the velocity vector magnitude for the new decoder optimised by the design tool now has a value of unity at the rear meeting one of Craven's goals. When examining the virtual microphone patterns of this decoder it appears that the addition of small higher order lobes for the centre and the front loudspeakers have contributed to the improved velocity vector response (figure 8-3 zooms in on the smaller higher order lobes of the centre and front left virtual microphones for both decoders).

Table 8.2 gives the objective values for the new decoder and the original Craven decoder. Overall the velocity vector angle objective (E_{LFAng}) and the energy vector angle objective (E_{HFAng}) are slightly worse in performance for the new decoder. However, the match between the vector angles around the sound stage has been improved meeting another of Craven's goals (highlighted by the better scores for $E_{AngMatch}$ in table 8.2).

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
New decoder	0.1226	0.0118	1.0872	28.3584	22.9281	48.8143	26.4282
Craven decoder	0.0753	0.0176	20.0059	28.8404	20.2021	46.9141	27.7488
	$E_{LFAngEv}$	$E_{HFAngEv}$	$E_{LFMagEv}$	$E_{HFMagEv}$	Total		
New decoder	0.1127	0.1594	0.1134	0.2751	128.3267		
Craven decoder	0.0053	0.1662	0.1226	0.2821	144.4647		

Table 8.2: Fitness function objective values for the Craven decoder and the decoder derived using Craven's decoder coefficients as a starting point.

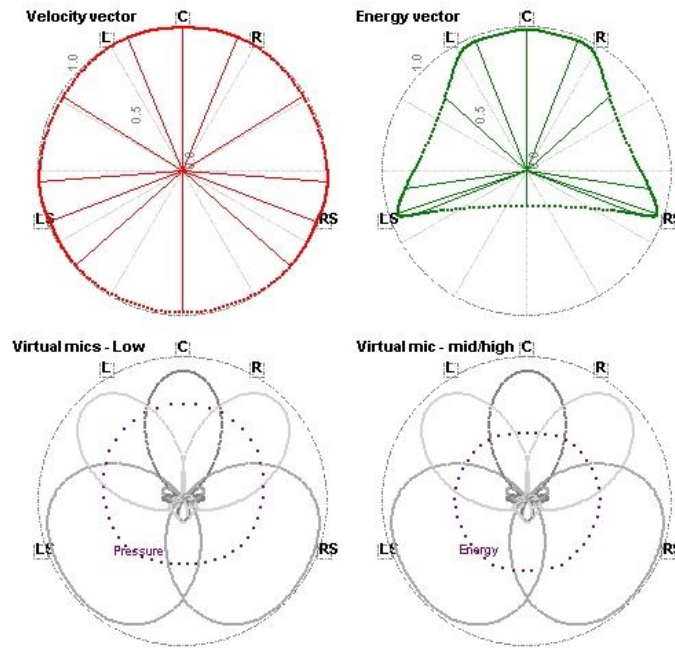


Figure 8-1: Performance plot of the best solution produced by the design tool when using Craven's decoder as a starting point.

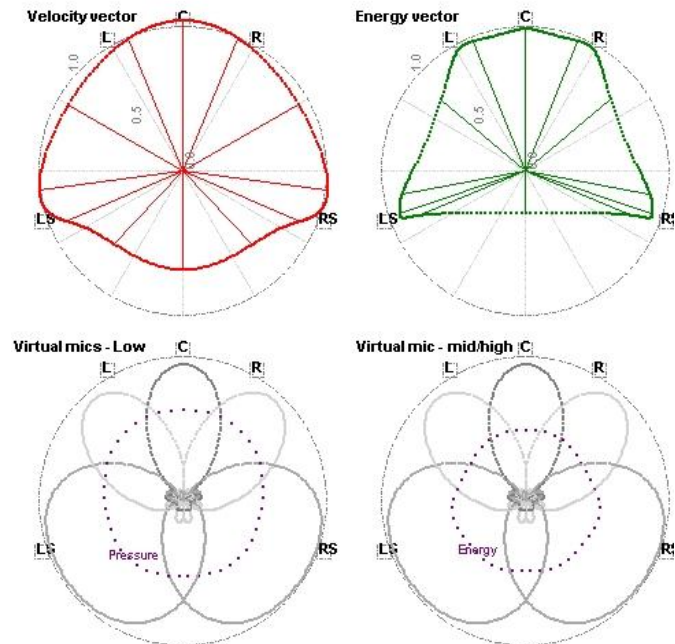


Figure 8-2: Performance plot of the decoder derived by Craven.

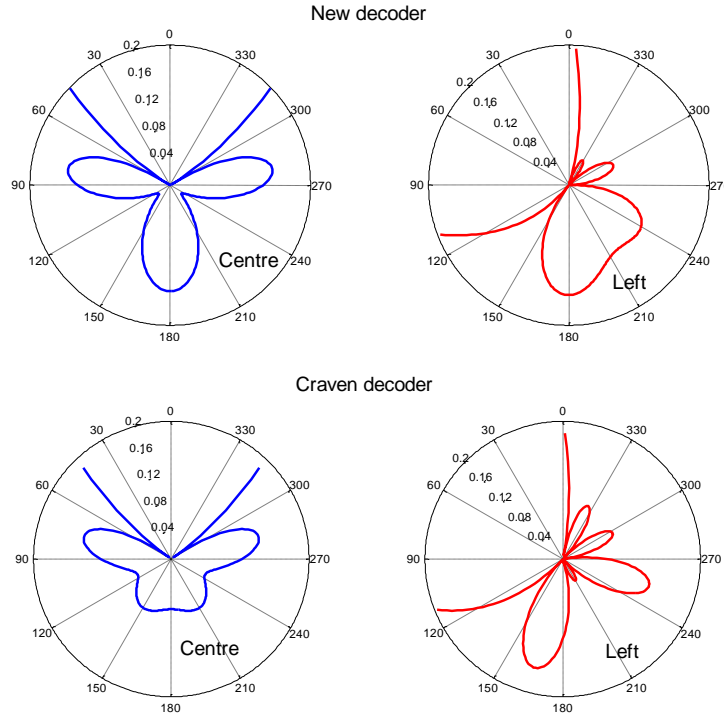


Figure 8-3: Zoomed in view of the centre and left front virtual microphones for the new decoder derived from Craven’s solution and the original Craven solution.

In addition to the improvements mentioned so far, the new decoder has reproduced energy that is more even by angle around the listener (see E_{HFVol} in table 8.2) and there has also been a small improvement for the energy vector magnitude objective (E_{HFMag}).

Although even performance was not part of Craven’s original design criteria, the even error objectives were still used in this application of the design tool (although with small importance weightings). Consequently, more even performance has been achieved for the new decoders for 3 out of 4 even error objectives ($E_{HFAngeV}$, E_{LFMag} , E_{HFMag}).

Overall, the decoder derived using the design tool has a total fitness score that is approximately 13% better than the Craven decoder (see table 8.2).

8.3 Further optimisation of the Poletti decoder

In both off-centre tests the decoder derived by Poletti came out as one of the best decoders and was broadly equivalent to the off-centre optimised decoder produced using the design tool (see chapter 6 and chapter 7). For that reason a further experimental application was carried out to see if the design tool could improve Poletti's decoder according to the methods developed in this research.

When deriving his decoder Poletti used a different optimisation approach (Poletti 2007). A least squares method was used for determining a set of loudspeaker weights which gave a good fit when matching the pressure generated by the loudspeakers at several points with the pressure generated by a plane wave at different angles. Although this method is different to the off-centre method developed in this research, the general aim is similar – to improve performance at distributed position in the listening area.

In his paper Poletti states that robust solutions possess a ‘double complementarity’ property where the sum of the loudspeaker weights (i.e. pressure) and the sum of the squared loudspeaker weights (i.e. energy) equal 1 for each angle around the 360° sound stage. In order to try and meet this as closely as possible, when using the design tool the low frequency volume objective and the mid/high frequency volume objective were given the most importance (see table 8.3).

E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
50.0	50.0	1.0	1.0	1.0	1.0	0.1
$E_{LFAngEv}$	$E_{HFAngEv}$	$E_{LFMagEv}$	$E_{HFMagEv}$			
0.0	0.0	0.0	0.0			

Table 8.3: Importance weights used when deriving a solution from Poletti's decoder

10 runs of the design tool were undertaken using Poletti's decoder as a starting point. Both the range-removal component and the off-centre optimisation component were switched on. Figure 8-4 displays the mean error of the velocity vector and energy vector magnitudes and angles for the best solution derived. The mean errors at each listening position take into account each

different source angle checked in the fitness function. Table ... shows the objective error values at the centre point. The performance of the original decoder by Poletti is included in both figure and table for comparison.

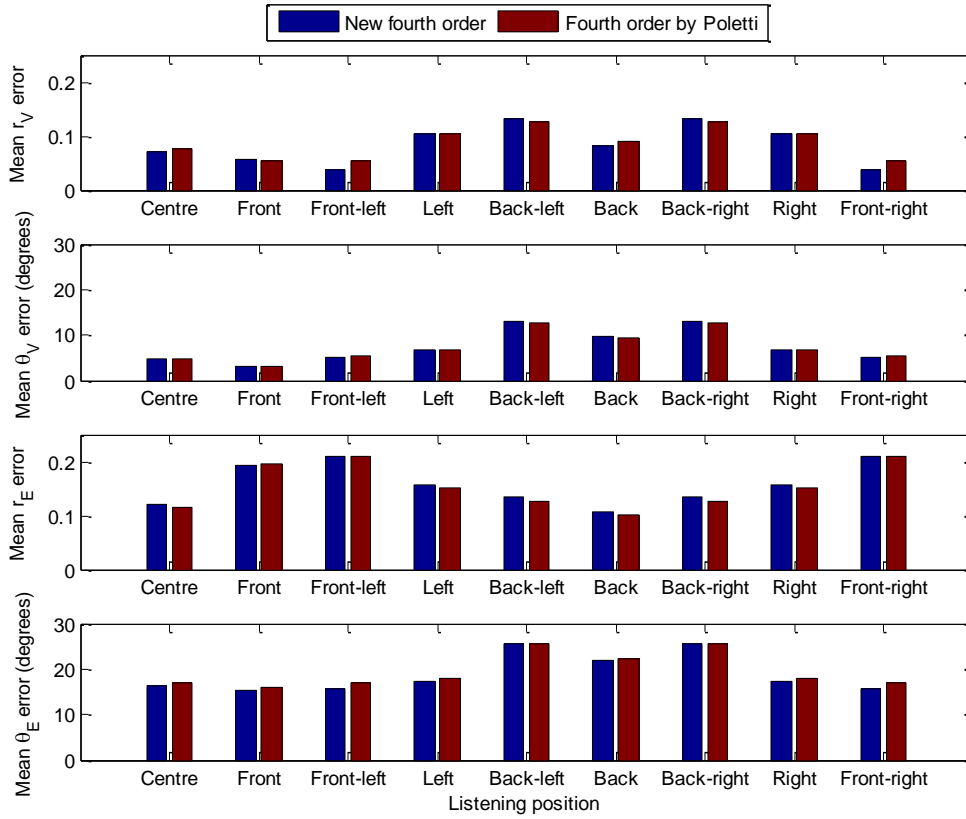


Figure 8-4: Mean velocity vector and energy vector errors for the new decoder derived from Poletti's decoder and the original Poletti's decoder at each position evaluated.

	E_{LFVol}	E_{HFVol}	E_{LFMag}	E_{HFMag}	E_{LFAng}	E_{HFAng}	$E_{AngMatch}$
New decoder	0.0000	0.1407	12.6975	22.2094	15.3485	51.6485	37.8370
Poletti decoder	0.0119	0.1506	14.1200	21.1418	15.4528	53.4801	39.2696

Table 8.4: Fitness function objective values for the decoder derived from Poletti and the original Poletti decoder.

Close examination of figure 8-4 shows that both decoders are very similar in terms of velocity vector and energy vector performance. The new decoder has better performance in some

positions but worse performance in others. However, the volume objective scores in table 8.4 show that the new decoder better meets the double complementarity criteria for a robust decoder.

Poletti's solution appears to be close to a local minimum in the domain of the fitness function developed in this work. Consequently, only a small improvement of approximately 1.5% was made in terms of total fitness. Nevertheless, an improvement has been made in terms meeting the double complementarity goal that was set.

8.4 Summary

This chapter has shown that the design tool is capable of taking existing decoders and improving them (according to the developed fitness function). Confidence gained in the design tool from the binaural tests and the listening tests implies that the solutions would perform well.

Chapter 9

Summary and Conclusions

9.1 Introduction

This research was motivated by the desire to produce improved playback over existing commercial surround sound loudspeaker arrays. The aim was to develop a tool capable of producing surround sound decoders that meet a variety of different design requirements for the central listening position and off-centre listening positions.

The thesis started with an introduction in chapter 1. Then, in Chapter 2, an extensive literature review was presented examining several topics relevant to this research: human auditory localisation, surround sound, optimisation using computer search algorithms and high performance computing. The review identified that Ambisonic surround sound decoder design for irregular loudspeaker arrays could be progressed further. In chapter 3, background theory was presented on the methods selected for use in this research.

The following section summarises the main contributions of this thesis from the work presented in chapters 4, 5, 6, 7, and 8.

9.2 Summary of the main contributions of this thesis

The major outcome of this work has been the development of a software-based decoder design tool that allows the user to produce decoders to desired criteria. To the best of the author's knowledge a decoder design tool of this sophistication has not been produced previously.

The tool encompasses a number of key components each of which can be controlled by selecting different options on the graphical user interface. The following remarks can be made based on the design tool and each of its components:

- A multi-objective fitness function was developed which more closely matches Gerzon's criteria for the Ambisonic system than in previous research work. The function comprises a new vector angle match objective and two improved reproduced volume objectives. The angle match objective ensures decoders are derived with a better match between the velocity vector angle and energy vector angle. The revised volume objectives are able to produce solutions with an even volume response by angle around the listener.
- The concept of "range-removal" was introduced to this application domain to resolve the problem of dominance in multi-objective fitness functions. Range-removal allows each of the fitness function objectives to have an equal impact in the search. A further concept known as 'importance' was introduced for logically biasing range-removed objectives. When used in combination, range-removal and importance enable good solutions to be derived by a search algorithm to user-defined criteria without *ad hoc* adjustment of objective weights.
- New fitness function objectives have been defined which enable decoders for irregular loudspeaker arrays to be derived with more even localisation performance by angle. Although there is a direct trade-off between good even performance and good overall performance, this trade-off can be balanced by using range-removal and importance.
- A novel concept was introduced where a decoder's theoretical localisation performance can be improved in specific areas of the sound stage by applying angle-dependent weightings to the fitness function objectives. The theoretical effectiveness of this concept was demonstrated by applying a weighting scheme according to the pattern of human spatial resolution determined in previous minimum audible angle experiments.
- A new method was developed to enable the localisation performance of decoders to be evaluated in off-centre listening positions. This method was incorporated into the fitness

function allowing the search to produce decoders with improved off-centre localisation performance.

- The design tool is able to produce Ambisonic decoders from first order to fourth order. Furthermore, these decoders can be frequency dependent or independent.
- The design tool optionally makes use of high performance computing hardware to accelerate the search process. The hardware enables more searches to be run in a period of time increasing the chance of finding a better solution in a fixed period of time. Alternatively, it makes the tool more interactive for the decoder designer.
- It is possible to start a search from an existing solution. Theoretical results show that when this is the case the tool is able to improve the theoretical performance of these solutions.

The tool provides the user with a wide range of options for fine tuning the performance of an irregular Ambisonic decoder to a level that has not been provided before. For example, a user might want to produce a frequency dependent decoder optimised for better mid/high frequency localisation performance in the front of the system; selecting the MAA checkbox (see element 3 in figure 4.8) and weighting the search in favour of the energy vector objectives would enable this. In contrast, the user might want to produce a decoder optimised for even localisation performance in off-centre listening positions. Selecting the off-centre checkbox (see element 3 in figure 4.8) and weighting in favour of the even error objectives would allow this.

Although the focus has been on Ambisonic decoders for the ITU layout with speaker at a constant radial distance, it should be stressed that the methods presented could be applied when designing Ambisonic decoders for any irregular surround sound loudspeaker layout. The current system could be expanded to account for this.

A further contribution of this thesis was the thorough investigation into the localisation performance of decoders developed using the design tool. The experiments conducted give a detailed insight into the performance of irregular 5-speaker Ambisonic decoders - not just decoders derived by the tool, but also some existing published decoders.

In the first experiment an extensive series of listening tests was conducted to investigate the performance of decoders developed for the ITU 5-speaker layout. In general, the results from the tests verify the design tool components developed in this thesis. This following summarises the main observations from the tests:

- The results of the listening tests suggest that the developed decoders perform at least as well as previously published decoders or better. For example, the standard fourth order decoder that was designed to equal or exceed the performance of existing published decoders was able to do so in the majority of cases.
- The even error optimisation component was, on balance, able to produce a decoder with even performance by angle. Even performance was most apparent for this decoder for the male speech source.
- The decoder generated using the MAA optimisation component gave excellent frontal performance at the cost of performance to the sides and rear for the majority of test cases.
- In off-centre listening positions the perception of phantom sound sources is problematic. The tests have provided an insight into the extent of these problems and shown that the off-centre optimised decoder was able to give significantly better performance than the non off-centre optimised decoders. This decoder was also able to equal (and arguably surpass) the performance of the only known existing off-centre optimised 5-speaker available in the literature.

- Generally localisation performance at the rear-sides of the ITU 5-speaker system was problematic for all decoders. However, sound images can still be perceived in this area from the optimal listening position and were improved for the best decoders.

In the second experiment measurements were made with a binaural microphone to further investigate the performance of the developed decoders. An auditory model was developed for assessing each decoder from the central measurement position and off-centre measurement positions. The following remarks can be made about these tests:

- The results are in broad agreement with the predicted performance further supporting the design criteria introduced in this thesis. Even performance was observed, albeit at a reduced overall level of performance, for the even error optimised decoder. The MAA does show some signs of improvement at the front in this test but performance at the sides and rear is degraded.
- All decoders analysed at the central position (apart from the default Soundfield decoder) are very good and the fitness function introduced in this work scores them all highly and quite similarly. This was reflected in the tests as the differences between the decoders are quite minimal.
- The off-centre measurements are in agreement with the predicted performance and also in agreement with the listening tests results. They show that the off-centre optimised decoder performs better in off-centre listening positions as desired at least equalling an existing off-centre optimised decoder.

The experiments in general support the predicted performance of the generated decoders and the effectiveness of the individual design tool components. However, further more extensive listening tests would be necessary to conclusively differentiate between the tested decoders. Some suggestions for further testing are included in the following chapter.

9.3 Conclusions

This thesis has presented a number of different methods that can be used in isolation or combination when improving surround sound playback over 5-speaker surround sound layouts with a constant radial distance from the listener. It has specifically focused on improving localisation performance of Ambisonic decoders for the ITU 5-speaker layout by developing optimisation criteria to be implemented with a search algorithm. The developed methods have been encapsulated in a fully functioning software tool.

In conclusion, the design tool presented in this thesis is able to produce improved Ambisonic decoders for the ITU 5-speaker layout. The performance of the developed decoders has been validated by binaural head tests and also listening tests. The tool offers decoder designers control when fine tuning the localisation performance of an Ambisonic system at central and non-central listening positions. This ultimately results in decoders that produce a more convincing illusion of surround sound for the listener.

Chapter 10

Future work

This research has uncovered a number of other potential avenues for future work. Some of the main ideas are included here.

Further in-depth evaluation of the design tool components - This project has largely been concerned with the development of a design tool. Extensive testing of 5-speaker decoders produced using the tool has been undertaken. However, to be completely conclusive, future tests could focus on a particular component of the tool in detail. For example, the tool could be used to produce a range of even error decoders with different levels of predicted even performance. A listening test could then be undertaken to investigate whether the predicted performance correlates with the listeners' responses.

Incorporating other perceptual measures into the design tool - In this work the focus has been on improving localisation. However, other perceptual factors are important for generating a convincing illusion of surround sound (e.g. envelopment). In the light of this, more objective measures could be developed and added to the fitness function. Range-removal and importance would allow the balance between these factors to be controlled.

Robust single decoder solution – This thesis investigated the performance of decoders designed for ideal systems with loudspeakers at an equal distance from the central listening point. In the home listening environment, however, loudspeakers are often positioned in a manner which is convenient. It would be interesting to investigate a single decoder solution that is robust to loudspeaker movement (i.e. one set of decoder coefficients for use over arbitrary loudspeaker layouts). It is envisaged that such a solution would be particularly useful for plug-and-play scenarios to allow the user to quickly experience Ambisonic surround.

Database of solutions – Create and build up a database of good solutions. The solutions could be derived using different combinations of components and importance weightings on the tool.

Further development of a design tool – Additional features could be added to the design tool. Some possible features include:

- The ability to optimise for different loudspeaker configurations (i.e. 7.1, 22.2).
- Allow the users to add any number of loudspeakers and position them on a graphical user interface. The user could also add listening positions. This data could then be used by the tool to produce a decoder suited to their personal situation.
- The ability to automatically and accurately identify the positions of the loudspeakers in relation to the listening point. This would allow the tool to further improve a decoder for a given layout.
- Decoder ‘audition’ function where the user can quickly assess the performance of the current best solution produced by the tool by listening to decoded audio.
- Incorporation of other perceptual measures for evaluating a system’s performance (as mentioned above).
- Testing of the tool for optimising decoders in a number of different listening scenarios (e.g. in-car listening, living room).

Finally, investigation is needed into the effect of the room on surround sound. Occasionally during the listening tests different artifacts were noticed when listening to panned sources (e.g.

elevated sources, sources sounding more distant than other). Unless anechoic, the characteristics of the listening room will be superimposed on the reproduced soundfield. This may affect localisation as well as other spatial attributes such as envelopment, localisation and distance perception. Further research is required to investigate the extent of the problem and also devise methods for compensating for this.

Appendix A

This appendix presents more efficient low and mid/high frequency volume objectives. Previously, in chapter 4, the volume at every angle was compared to the volume at all other angles resulting in N^2 iterations for each volume objective (where N is the number of angles checked in the fitness function).

In order to offer a more efficient (but equally effective) means of measuring volume equality around the 360° sound stage a different objective design was developed. The alternative method involves computing a running standard deviation of the volume error at each angle i.e.

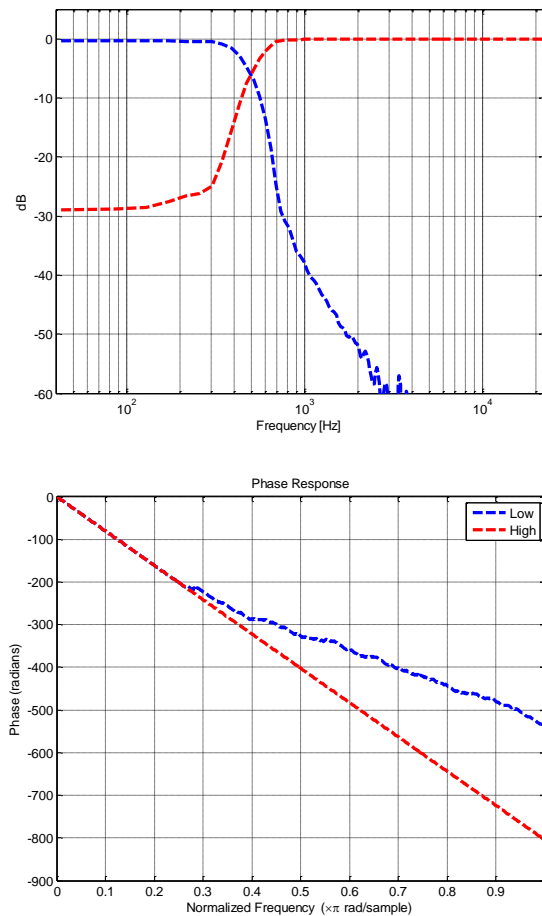
$$E_{LFVol} = \frac{1}{N-1} \left[\sum_{i=0}^{N-1} P_i^2 - \frac{1}{N} \left(\sum_{i=0}^{N-1} P_i \right)^2 \right]$$

$$E_{HFVol} = \frac{1}{N-1} \left[\sum_{i=0}^{N-1} E_i^2 - \frac{1}{N} \left(\sum_{i=0}^{N-1} E_i \right)^2 \right]$$

Where N is the number of angles to check, P is the pressure, E is the energy. The running standard deviation algorithm was taken from (Smith 1998). The advantage of using the running standard deviation (rather than the direct method) is the standard deviation can be determined from a single pass through the data. The direct method on the other hand has to pass through the data set twice to first calculate the mean.

Appendix B

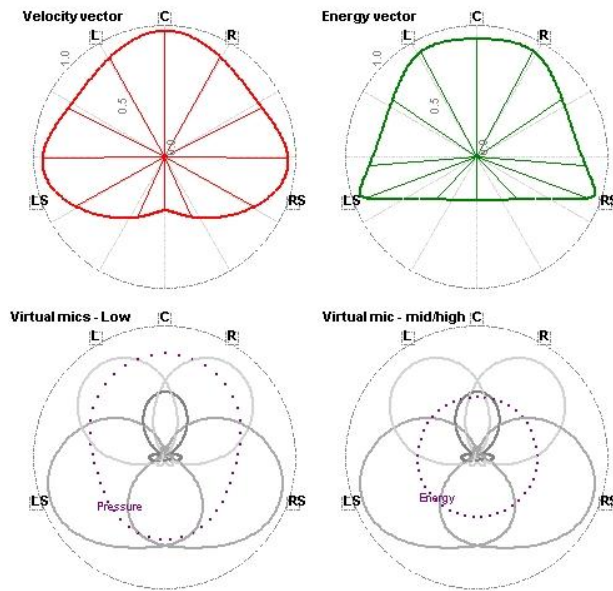
In the listening test and the binaural test a first order frequency dependent decoder was evaluated from the centre listening position (i.e. Decoder 2 in Chapter 6 and Chapter 7). When implementing this decoder two linear phase FIR filters were used to divide the X and Y encoded audio signals into two separate frequency regions with a cross-over frequency of 500 Hz. The following figures show the magnitude and phase response of both filters.



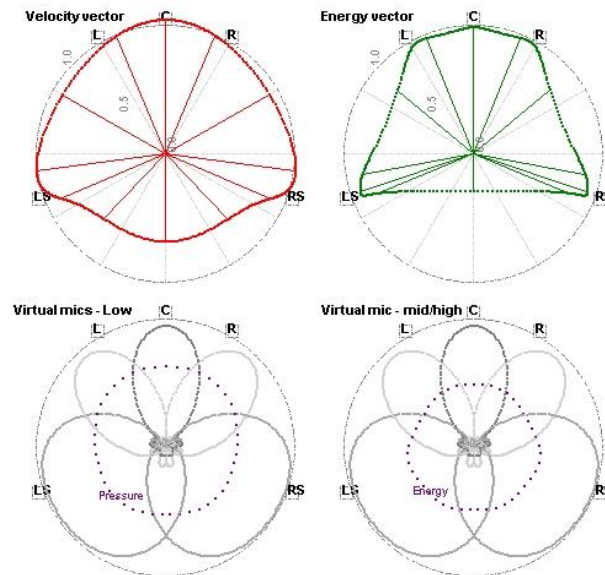
The high frequency filter is phase linear across all frequencies whereas the low frequency filter is phase linear up to approximately 6 kHz. Linear phase filters are recommended when implementing frequency dependent Ambisonic decoders (Lee 2005).

Decoders developed by other authors were also evaluated in the listening tests and the binaural tests. These included fourth order decoders developed by Wiggins (Wiggins 2007), Craven (Craven 2003) and Poletti (Poletti 2007). The performance plots for each decoder are included below.

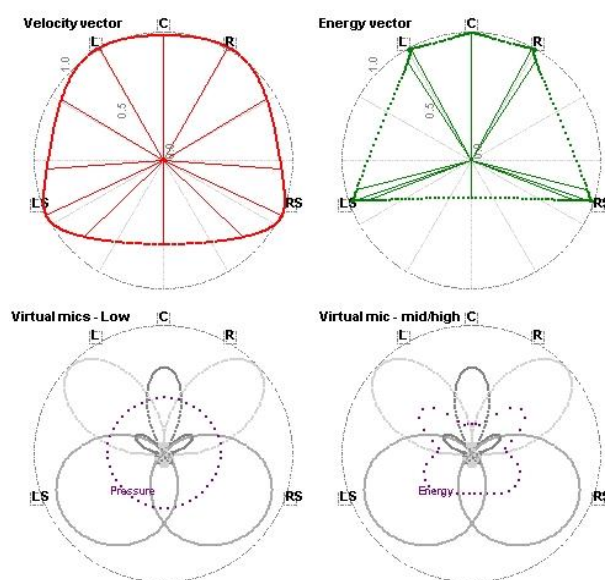
Decoder developed by Wiggins



Decoder developed by Craven



Decoder developed by Poletti



Appendix C

Tests of Within-Subjects Effects							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Sound	Greenhouse-Geisser	11912.531	2.133	5584.590	11.549	.000	.044
Sound * Subject	Greenhouse-Geisser	15490.977	27.730	558.628	1.155	.269	.056
Error(Sound)	Greenhouse-Geisser	259929.825	537.543	483.552			

Tests of Between-Subjects Effects							
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	
Intercept	48963.409	1	48963.409	184.257	.000	.422	
Subject	6738.346	13	518.334	1.951	.025	.091	
Error	66964.912	252	265.734				

Table 1: 2-way mixed design ANOVA showing the significance sound and subject and their interaction.

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Sound	Greenhouse-Geisser	6864.912	1.671	4109.298	8.501	.001	.031
Error(Sound)	Greenhouse-Geisser	214001.754	442.704	483.397			

Table 2: 1-way within-subjects ANOVA showing the significance of the sound source (original data).

Source		Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Sound	Greenhouse-Geisser	614.286	1.886	325.648	9.854	.000	.036
Error(Sound)	Greenhouse-Geisser	16519.048	499.882	33.046			

Table 3: 1-way within-subjects ANOVA showing the significance of the sound source (reversal corrected data).

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Corrected Model	LF (original)	14734.211 ^a	13	1133.401	3.366	.000	.148
	HF (original)	2975.940 ^b	13	228.918	.306	.991	.016
	Speech (original)	2093.233 ^c	13	161.018	1.432	.145	.069
	LF (corrected)	839.474 ^d	13	64.575	1.544	.102	.074
	HF (corrected)	744.361 ^e	13	57.259	1.528	.107	.073
	Speech (corrected)	842.105 ^f	13	64.777	3.321	.000	.146
Intercept	LF (original)	17702.632	1	17702.632	52.568	.000	.173
	HF (original)	27813.534	1	27813.534	37.122	.000	.128
	Speech (original)	2780.451	1	2780.451	24.736	.000	.089
	LF (corrected)	5323.684	1	5323.684	127.322	.000	.336
	HF (corrected)	5413.534	1	5413.534	144.482	.000	.364
	Speech (corrected)	1842.105	1	1842.105	94.433	.000	.273
Subject	LF (original)	14734.211	13	1133.401	3.366	.000	.148
	HF (original)	2975.940	13	228.918	.306	.991	.016
	Speech (original)	2093.233	13	161.018	1.432	.145	.069
	LF (corrected)	839.474	13	64.575	1.544	.102	.074
	HF (corrected)	744.361	13	57.259	1.528	.107	.073
	Speech (corrected)	842.105	13	64.777	3.321	.000	.146
Error	LF (original)	84863.158	252	336.759			
	HF (original)	188810.526	252	749.248			
	Speech (original)	28326.316	252	112.406			
	LF (corrected)	10536.842	252	41.813			
	HF (corrected)	9442.105	252	37.469			
	Speech (corrected)	4915.789	252	19.507			
Total	LF (original)	117300.000	266				
	HF (original)	219600.000	266				
	Speech (original)	33200.000	266				

continued ...

	LF (corrected)	16700.000	266
	HF (corrected)	15600.000	266
	Speech (corrected)	7600.000	266
Corrected Total	LF (original)	99597.368	265
	HF (original)	191786.466	265
	Speech (original)	30419.549	265
	LF (corrected)	11376.316	265
	HF (corrected)	10186.466	265
	Speech (corrected)	5757.895	265

a. R Squared = .148 (Adjusted R Squared = .104)

b. R Squared = .016 (Adjusted R Squared = -.035)

c. R Squared = .069 (Adjusted R Squared = .021)

d. R Squared = .074 (Adjusted R Squared = .026)

e. R Squared = .073 (Adjusted R Squared = .025)

f. R Squared = .146 (Adjusted R Squared = .102)

Table 4: 1-way between-subjects ANOVA for all sound sources showing the significance of subjects (original data and reversal corrected).

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	192135.807	6.117	31412.272	24.210	.000	.132
Error(Decoder)	Greenhouse-Geisser	1261843.359	972.537	1297.476			
Sound	Greenhouse-Geisser	226156.302	1.708	132404.368	34.383	.000	.178
Error(Sound)	Greenhouse-Geisser	1045827.031	271.584	3850.848			
Decoder * Sound	Greenhouse-Geisser	30242.865	10.616	2848.722	2.269	.011	.014

Table 5: 2-way within subjects ANOVA showing the significance of the sound source and decoder and their interaction (original data)

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	57378.229	5.093	11266.467	51.440	.000	.244
Error(Decoder)	Greenhouse-Geisser	177355.104	809.760	219.022			
Sound	Sphericity Assumed	4621.927	2	2310.964	18.552	.000	.104
Error(Sound)	Sphericity Assumed	39611.406	318	124.564			
Decoder * Sound	Greenhouse-Geisser	3367.240	11.991	280.816	2.133	.013	.013
Error(Decoder*Sound)	Greenhouse-Geisser	250999.427	1906.557	131.651			

Table 6: 2-way within subjects ANOVA showing the significance of the sound source and decoder and their interaction (reversal corrected data).

Source	Sphericity Correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	68312.500	5.744	11893.290	8.358	.000	.050
Error(Decoder)	Greenhouse-Geisser	1299487.500	913.262	1422.908			

Table 7: 1-way within-subjects ANOVA showing the significance of decoder when taking into account the original data from the low frequency noise test.

1-way within-subjects ANOVA showing the significance of decoder for low frequency noise data (corrected)

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	24758.672	5.442	4549.935	27.552	.000	.148
Error(Decoder)	Greenhouse-Geisser	142878.828	865.205	165.139			

Table 8: 1-way within-subjects ANOVA showing the significance of decoder when taking into account the reversal corrected data from the low frequency noise test.

1-way within-subjects ANOVA showing the significance of decoder for mid/high frequency noise data (original)

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	115849.922	6.184	18733.495	11.237	.000	.066
Error(Decoder)	Greenhouse-Geisser	1639312.578	983.273	1667.200			

Table 9: 1-way within-subjects ANOVA showing the significance of decoder when taking into account the original data from the mid/high frequency noise test.

1-way within-subjects ANOVA showing the significance of decoder for mid/high frequency noise data (corrected)

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	15315.547	6.123	2501.436	15.408	.000	.088
Error(Decoder)	Greenhouse-Geisser	158046.953	973.510	162.348			

Table 10: 1-way within-subjects ANOVA showing the significance of decoder when taking into account the reversal corrected data from the mid/high frequency noise test.

1-way within-subjects ANOVA showing the significance of decoder for male speech data (original)

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	38216.250	5.099	7494.858	13.745	.000	.080
Error(Decoder)	Greenhouse-Geisser	442083.750	810.740	545.284			

Table 11: 1-way within-subjects ANOVA showing the significance of decoder when taking into account the original data from the male speech test.

1-way within-subjects ANOVA showing the significance of decoder for male speech data (corrected)

Source	Sphericity correction	Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Greenhouse-Geisser	20671.250	5.680	3639.603	25.793	.000	.140
Error(Decoder)	Greenhouse-Geisser	127428.750	903.046	141.110			

Table 12: 1-way within-subjects ANOVA showing the significance of decoder when taking into account the corrected data from the male speech test.

Source		Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Sound	Sphericity Assumed	101803.040	2	50901.520	4.317	.015	.049
Error(Sound)	Sphericity Assumed	1957343.458	166	11791.226			
Decoder	Sphericity Assumed	1485732.782	4	371433.195	35.827	.000	.302
Error(Decoder)	Sphericity Assumed	3442007.080	332	10367.491			
Position	Greenhouse-Geisser	7610302.418	2.633	2889807.697	186.705	.000	.692
Error(Position)	Greenhouse-Geisser	3383172.903	218.580	15477.939			
Sound * Decoder	Greenhouse-Geisser	62219.143	6.501	9570.428	.775	.599	.009
Error(Sound*Decoder)	Greenhouse-Geisser	6659357.382	539.599	12341.319			
Sound * Position	Greenhouse-Geisser	448487.463	4.630	96855.223	5.499	.000	.062
Error(Sound*Position)	Greenhouse-Geisser	6768729.441	384.331	17611.721			
Decoder * Position	Greenhouse-Geisser	873176.816	8.458	103235.341	7.246	.000	.080
Error(Decoder*Position)	Greenhouse-Geisser	1.000E7	702.024	14247.548			
Sound * Decoder * Position	Greenhouse-Geisser	209374.655	15.185	13788.222	.909	.554	.011
Error(Sound*Decoder*Position)	Greenhouse-Geisser	1.911E7	1260.358	15165.992			

Table 13: Multi-way within-subjects ANOVA showing the significance of decoder, sound source and listening position.

Source		Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Sphericity Assumed	320392.650	4	80098.162	10.816	.000	.115
Error(Decoder)	Sphericity Assumed	2458738.512	332	7405.839			
Position	Greenhouse-Geisser	2351974.980	2.881	816326.549	56.013	.000	.403
Error(Position)	Greenhouse-Geisser	3485141.174	239.137	14573.824			
Decoder * Position	Greenhouse-Geisser	210178.992	11.606	18108.749	2.000	.023	.024
Error(Decoder*Position)	Greenhouse-Geisser	8723765.552	963.339	9055.763			

Table 14: 2-way within-subjects ANOVA for the low frequency noise source.

Source		Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Sphericity Assumed	477994.330	4	119498.583	12.115	.000	.127
Error(Decoder)	Sphericity Assumed	3274809.643	332	9863.884			
Position	Greenhouse-Geisser	2507112.503	2.585	969758.416	60.036	.000	.420
Error(Position)	Greenhouse-Geisser	3466110.245	214.580	16153.031			
Decoder * Position	Greenhouse-Geisser	326527.047	8.114	40240.114	3.103	.002	.036
Error(Decoder*Position)	Greenhouse-Geisser	8732960.222	673.501	12966.520			

Table 15: 2-way within-subjects ANOVA for the mid/high frequency noise source.

Source		Type III Sum of Squares	df	Mean Square	F	Sig. <i>p</i>	Partial Eta Squared
Decoder	Sphericity Assumed	749564.945	4	187391.236	14.244	.000	.146
Error(Decoder)	Sphericity Assumed	4367816.307	332	13156.073			
Position	Greenhouse-Geisser	3199702.398	2.257	1417535.709	82.975	.000	.500
Error(Position)	Greenhouse-Geisser	3200650.925	187.350	17083.806			
Decoder * Position	Greenhouse-Geisser	545845.433	7.389	73876.812	3.886	.000	.045
Error(Decoder*Position)	Greenhouse-Geisser	1.166E7	613.253	19013.319			

Table 16: 2-way within-subjects ANOVA for the male speech source.

Bibliography

- Aarts, R.M. Enlarging the Sweet Spot for Stereophony by Time/Intensity Trading. Presented at the 94th Audio Engineering Society Convention, Berlin, Germany. 1993.
- Abhayapala, T. & Ward, D. Theory and design of high order sound field microphones using spherical microphone array. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. pp.1949-1952. 2002.
- AES. A Symposium on Multichannel Audio for Radio Broadcasters. *Journal of the Audio Engineering Society*, 52(10), pp. 1066-1071, 2004.
- Akeroyd, M., MRC Institute of Hearing Research. Available at: <http://www.ihr.mrc.ac.uk/products/index.php?page=matlab> [Accessed July 8, 2009].
- Anderson, D. & Fedak, G. The Computational and Storage Potential of Volunteer Computing. *Proceedings of the 6th IEEE International Symposium on Cluster Computing and the Grid*, pp. 73-80, 2006.
- Aurora, 2008. AURORA Plug-ins. Available at: http://www.aurora-plugins.com/Aurora_XP/index.htm [Accessed December 4, 2008].
- Back, T. & Schwefel, H. An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation*, 1(1), pp. 1-23, 1993.
- Bamford, J.S. & Vanderkooy, J. Ambisonic Sound for Us. Presented at the 99th Audio Engineering Society Convention, New York, USA, 1995.
- Barron, M. & Marshall, A.H. Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *Journal of Sound and Vibration*, 77, pp.211-32, 1981.
- Bates, E., Kearney, G., Boland, F., Furlong, D. Monophonic Source Localization for a Distributed Audience in a Small Concert Hall. Presented at the 10th International Conference on Digital Audio Effects, Bordeaux, France, 2007.
- Battiti, R. & Protasi, M., 1997. Reactive search, a history-based heuristic for MAX-SAT. *ACM Journal of Experimental Algorithmics*, 2. Available at:

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.5954> [Accessed November 26, 2008].

- Bauer, B.B. Phasor Analysis of Some Stereophonic Phenomena. *The Journal of the Acoustical Society of America*, 33(11), pp.1536-1539, 1961.
- Beckinger, M. & Brix, S. An Efficient Method to Generate Particle Sounds in Wave Field Synthesis. Presented at the 125th Audio Engineering Society Convention, San Francisco, USA, 2008.
- Benjamin, E. An Experimental Verification of Localization in Two-Channel Stereo. Presented at the 121st Audio Engineering Society Convention, San Francisco, USA, 2006.
- Benjamin, E. & Brown, P. The Effect of Head Diffraction on Stereo Localisation in the Mid-Frequency Range. Presented at the 122nd Audio Engineering Society Convention, Vienna, Austria, 2007.
- Benjamin, E., Lee, R. & Hellar, A. Localization in Horizontal-Only Ambisonic Systems. Presented at the 121st Audio Engineering Society Convention, San Francisco, USA, 2006.
- Bennett, J.C., Barker, K. & Edeko, F.O. A New Approach to the Assessment of Stereophonic Sound System Performance. *Journal of the Audio Engineering Society*, 33(5), pp.314-321, 1985.
- Bentley, P.J. & Wakefield, J.P. Finding Acceptable Solutions in the Pareto-Optimal Range using Multiobjective Genetic Algorithms. In P. K. Chawdhry, R. Roy, & R. K. Pant, eds. *Soft Computing in Engineering Design and Manufacturing*. Springer Verlag London Limited, pp. 231-240, 1998.
- Berkhout, A.J. A Wavefield Approach to Multichannel Sound. Presented at the 104th Audio Engineering Society Convention, Amsterdam, The Netherlands, 1998.
- Berkhout, A.J., Vries, D.D. & Vogel, P. Wave Front Synthesis: A New Direction in Electroacoustics. Presented at the 93rd Audio Engineering Society Convention, San Francisco, USA, 1992.
- Bernfeld, B. Simple Equations For Multichannel Stereophonic Sound Localization. *Journal of the Audio Engineering Society*, 23(7), pp.553-557, 1975.
- Bernstein, L.R., van de Par, S. & Trahiotis, C. The normalized interaural correlation:

- Accounting for NoS pi thresholds obtained with Gaussian and ``low-noise" masking noise. *The Journal of the Acoustical Society of America*, 106(2), pp.870-876, 1999.
- Betlehem, T. & Abhayapala, T.D. Theory and design of sound field reproduction in reverberant rooms. *The Journal of the Acoustical Society of America*, 117(4), pp.2100-2111, 2005.
- Biundo, S. & Fox, M. Recent Advances in AI Planning, Proceedings of the 5th European Conference on Planning, Springer, 1999.
- Blauert, J. *Communication Acoustics* 1st ed., Springer, 2005.
- Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, 2001.
- Blauert, J. & Cobben, W. Some Consideration of Binaural Cross-Correlation Analysis. *Acustica*, 39, pp.96-104, 1978.
- Bloom, J.P. Creating Source Elevation Illusions by Spectral Manipulation. *Journal of the Audio Engineering Society*, 25(9), pp.560-565, 1977.
- Blumlein, A.D., 1937. Sound-transmission, sound-recording, and sound-reproducing system. Available at: <http://www.freepatentsonline.com/2093540.html> [Accessed January 7, 2009].
- Boar, K.D. Stereophonic Sound Production. *Phillips Technical Review*, 5, pp.107-144, 1940.
- Boslaugh, S. & Watters, P. *Statistics in a Nutshell: A Desktop Quick Reference (In a Nutshell)*, O'Reilly Media, Inc, 2008.
- Braasch, J. A Binaural Model to Predict Position and Extension of Spatial Images Created with Standard Sound Recording Techniques. Presented at the 119th Audio Engineering Society Convention, New York, USA, 2005.
- Brungart, D.S. & Rabinowitz, W.M. Auditory Localization of Nearby Sources. Head Related Transfer Functions. *The Journal of the Acoustical Society of America*, 106(3), pp.1465-1479, 1999.
- Chandler, D.W., Grantham, D.W. & Leek, M.R. Effects of Uncertainty on auditory spatial resolution in the horizontal plane. *Acustica*, 91(3), pp.513-525, 2005.

- Chapman, M. New Dimensions for Ambisonics. Presented at the 124th Audio Engineering Society Convention, Amsterdam, The Netherlands, 2008.
- Cheng, C.I. & Wakefield, G.H. Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. *Journal of the Audio Engineering Society*, 49(4), pp.231-249, 2001.
- CIPIC, 2004. CIPIC Interface Laboratory: HRTF Database. *The CIPIC HRTF Database*. Available at: http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm [Accessed January 10, 2009].
- Clark, H.A.M., Dutton, G.F. & Vanderlyn, P.B. The 'Stereosonic' Recording and Reproducing System: A Two-Channel Systems for Domestic Tape Records. *Journal of the Audio Engineering Society*, 6(2), pp.102-117, 1958.
- Clearspeed, 2008. ClearSpeed - Home. Available at: <http://www.clearspeed.com/> [Accessed November 26, 2008].
- Cooper, D.H. & Bauck, J.L. Prospects for Transaural Recording. Presented at the 85th Audio Engineering Society Convention, Los Angeles, USA, 1988.
- Cooper, D.H. & Shiga, T. Discrete-Matrix Multichannel Stereo. *Journal of the Audio Engineering Society*, 20(5), pp.346-360, 1972.
- Corey, J. & Woszczyk, W. Localization of Lateral Phantom Images in a 5-Channel System with and without Simulated Early Reflections. Presented at the 113th Audio Engineering Society Convention, Los Angeles, USA, 2002.
- Cotterell, P. *On the Theory of the Second-Order Soundfield Microphone*. PhD. University of Reading, 2002.
- Craven, P.G. Continuous Surround Panning for 5-Speaker Reproduction. Presented at the 24th International Audio Engineering Society Conference, Banff, Canada, 2003.
- Craven, P.G. & Gerzon, M.A. Coincident microphone simulation covering three dimensional space and yielding various directional outputs. US Patent 4042779, 1977.
- Damaske, P. & Ando, P. Interaural Crosscorrelation for Multichannel Loudspeaker

- Reproduction. *Acustica*, 27, pp.232-238, 1972.
- Daniel, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD. University of Paris, 2001.
- Daniel, J. & Moreau, S. Further Study of Sound Field Coding with Higher Order Ambisonics. Presented at the 116th Audio Engineering Society Convention, Berlin, Germany, 2004.
- Daniel, J., Nicol, R. & Moreau, S. Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. Presented at the 114th Audio Engineering Society Convention, Amsterdam, The Netherlands, 2003.
- Daniel, J., Rault, J. & Polack, J. Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. Presented at the 105th Audio Engineering Society Convention, San Francisco, USA, 1998.
- Davis, M.F. History of Spatial Coding. *Journal of the Audio Engineering Society*, 51(6), pp.554-569, 2003.
- Dell'amico, M., Lodi, A. & Maffioli, F. Solution of the Cumulative Assignment Problem With a Well-Structured TabuSearch Method. *Journal of Heuristics*, 5(2), pp.123-143, 1999.
- Dobrucki, A. & Plaskota, P. Head-Related Transfer Function Calculation Using Boundary Element Method. Presented at the 122th Audio Engineering Society Convention, Vienna, Austria, 2007.
- Dolby, Surround Sound: Past, Present and Future. Available at: [http://www.dolby.com/uploadedFiles/English_\(US\)/Professional/Technical_Library/Technologies/Dolby_Surround/2_Surround_Past.Present.pdf](http://www.dolby.com/uploadedFiles/English_(US)/Professional/Technical_Library/Technologies/Dolby_Surround/2_Surround_Past.Present.pdf) [Accessed November 26, 2008].
- El-Rewini, H. & Abd-El-Barr, M. *Advanced Computer Architecture and Parallel Processing (Wiley Series on Parallel and Distributed Computing)* 1st ed., Wiley-Interscience, 2005.
- Farina, A. Advancements in Impulse Response Measurements by Sine Sweeps. Presented at the 122th Audio Engineering Society Convention, Vienna, Austria, 2007.
- Farina, A. Simultaneous Measurement of Impulse Response and Distortion with a Swept

- Sine. Presented at the 108th Audio Engineering Society Convention, Paris, France, 2000.
- Feddersen, W.E. Localization of High-Frequency Tones. *The Journal of the Acoustical Society of America*, 29(9), pp.988-991, 1957.
- Gardner, M.B. & Gardner, R.S. Problem of Localization in the Median Plane: Effect of Pinnae Cavity Occlusion. *The Journal of the Acoustical Society of America*, 53(2), pp.400-408, 1973.
- Gaspero, L.D. & Schaerf, A. A composite-neighborhood tabu search approach to the traveling tournament problem. *Journal of Heuristics*, 13(2), pp.189-207, 2007.
- Gaston, L. Methods for Sharing Stereo and Multichannel Recordings Among Planetariums. Presented at the 124th Audio Engineering Society Convention Amsterdam, The Netherlands, 2008.
- Gendreau, M., Hertz, A. & Laporte, G. A Tabu Search Heuristic for the Vehicle Routing Problem. *Manage. Sci.*, 40(10), pp.1276-1290, 1994.
- Gerzon, M.A. Ambisonics in Multichannel Broadcasting and Video. *Journal of the Audio Engineering Society*, 33(11), pp.859-871, 1985.
- Gerzon, M.A. General Metatheory of Auditory Localisation. Presented at the 92nd Audio Engineering Society Convention, Vienna, Austria, 1992.
- Gerzon, M.A. Optimum Reproduction Matrices for Multispeaker Stereo. *Journal of the Audio Engineering Society*, 40(7/8), pp.571-589, 1992.
- Gerzon, M.A. Practical Periphony: The Reproduction of Full-Sphere Sound. Presented at the 65th Audio Engineering Society Convention, London, England, 1980.
- Gerzon, M.A. Surround-sound psychoacoustics. *Wireless World*, 80, pp.483-486, 1974.
- Gerzon, M.A. & Barton, G.J. Ambisonic Decoders for HDTV. Presented at the 92nd Audio Engineering Society Convention Vienna, Austria., 1992.
- Gerzon, M.A. & Barton, G.J. Surround Sound Apparatus. US Patent 5757527, 1998.
- Glasgal, R. Ambiophonics: Achieving Psychological Realism in Music Recording and Reproduction. Presented at the 111th Audio Engineering Society Convention, New York, USA, 2001.

- Glover, F. Tabu Search - Part 1. *ORSA Journal of Computing*, 1(3), pp.190-206, 1989.
- Glover, F. Tabu Search - Part 2. *ORSA Journal of Computing*, 2(1), pp.4-32, 1990.
- Grantham, D.W., Hornsby, B.W.Y. & Erpenbeck, E.A. Auditory Spatial Resolution in Horizontal, Vertical, and Diagonal Planes. *The Journal of the Acoustical Society of America*, 114(2), pp.1009-1022, 2003.
- Griesinger, D. Objective Measures of Spaciousness and Envelopment. Presented at the 16th International Audio Engineering Society Conference, Finland, 1999.
- Guastavino, C., Larcher, V., Catusseaus G. & Boussards, P. Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo. Presented at the 13th International Conference on Auditory Display, Montreal, Canada, 2007.
- Hamasaki, K., Nishiguchi, T. & Ono, K. Advanced Multichannel Audio Systems with Superior Impression of Presence and Reality. Presented at the 116th Audio Engineering Society Convention, Berlin, Germany, 2004.
- Hammer, K.E. & Snow, W.B. *Binaural Transmission System at the Academy in Philadelphia* J. Allen, ed., Bell Telephone Laboratories, 1932.
- Han, H. Measuring a Dummy Head in Search of Pinna Cues. *Journal of the Audio Engineering Society*, 42(1/2), pp.15-37, 1994.
- Hartmann, W.M. On the Minimum Audible Angle - A Descision Theory Approach. *The Journal of the Acoustical Society of America*, 85(5), pp.2031-2041, 1989.
- Hertz, A., Taillard, E. & de Werra, D, *A Tutorial on Tabu Search*, In the Proceedings of Giornate di Lavoro, 1995.
- Holman, T. *5.1 Surround Sound: Up and Running*. Focal Press, 2000.
- Home Audio Division. CEA: Home Audio Division - Home Audio Division. Available at: <http://www.ce.org/Membership/Divisions/91.asp> [Accessed November 23, 2008].
- Howard, D.M. & Angus, J. *Acoustics and Psychoacoustics*, Focal Press, 2001.
- Ibbotson, J. Hollywood sound for Cricklewood Money. Presented at the 22nd UK Audio Engineering Society Conference (*Illusions in Sound: the application of*

- psychoacoustics to audio*), Cambridge, UK, 2007.
- Inanaga, K., Yamada, Y. & Koizumi, H. Headphone System with out-of-head Localisation Applying Dynamic HRTF (Head Related Transfer Function). Paris, France, 1995.
- ISO. *Acoustics - Measurement of the Reverberation Time of Rooms with Reference to other Acoustical Parameters*, Geneva, Switzerland: International Organisation for Standardisation, 1997.
- ITU. Multichannel stereophonic sound system with and without accompanying picture, 1994.
- Jackson, P., Dewhirst, M., Conetta, R., Zielinski, S., Rumsey, F., Meares, D., Bech, S. & George, S. QESTRAL (Part 3): system and metrics for spatial quality prediction. Presented at the 125th Audio Engineering Society Convention, San Francisco, USA, 2008.
- Jeppesen, J. & Moller, H. Cues for Localisation in the Horizontal Plane. Presented at the 118th Audio Engineering Society Convention, Barcelona, Spain, 2005.
- Jot, J., Larcher, V. & Pernaux, J. A Comparative Study of 3-D Audio Encoding and Rendering Techniques. Presented at the 16th International Audio Engineering Society Conference, Finland, 1999.
- Jot, J., Larcher, V. & Warusfel, O. Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony. Presented at the 98th Audio Engineering Society Convention, Paris, France, 1995.
- Kato, M., Uematsu, H., Kashino, M. & Hirahara, T. The Effect of Head Motion on the Accuracy of Sound Localization. *Acoustical Science and Technology*, 24(5), pp.315-317, 2003.
- Katz, B.F.G. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *The Journal of the Acoustical Society of America*, 110(5), pp.2440-2448, 2001.
- Katz, B.F.G.. Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements. *The Journal of the Acoustical Society of America*, 110(5), pp.2449-2455, 2001.
- Kearney, G., Boland, F., Furlong, D., Bates, E. A Comparative Study of the Performance

- of Spatialization Techniques for a Distributed Audience in a Concert Hall Environment. Presented at the 31st International Audio Engineering Society Conference, London, England, 2007.
- Khanna. gravity.phy.umassd.edu. Available at: <http://gravity.phy.umassd.edu/ps3.html> [Accessed November 26, 2008].
- Kirkeby, O. & Nelson, P.A. The "Stereo Dipole" - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers. *Journal of the Audio Engineering Society*, 46(5), pp.387-395, 1997.
- Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. Optimization by simulated annealing. *Science*, 220, pp.671-680, 1983.
- Kistler, D.J. & Wightman, F.L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3), pp.1637-1647, 1992.
- Langendijk, E.H.A. & Bronkhorst, A.W. Contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America*, 112(4), pp.1583-1596, 2002.
- Leakey, D.M. Some Measurements on the Effects of Interchannel Intensity and Time Differences in Two Channel Systems. *The Journal of the Acoustical Society of America*, 31(7), pp.977-986, 1959.
- Lee, R.. Shelf Filters for Ambisonic Decoders, Available at: <http://www.ambisonia.com/Members/ricardo/shelfs.zip/view>, 2005
- Lee, R. & Hellar, A.J. Ambisonic Localisation - Part 2. Presented at the 14th International Congress on Sound and Vibration, Cairns, Australia, 2007.
- Litovsky, R.Y. & Colburn, S.H. The Precedence Effect. *The Journal of the Acoustical Society of America*, 106(4), pp.1633-1654, 1999.
- Mac Cabe, C.J. & Furlong, D.J. Virtual Imaging Capabilities of Surround Sound Systems. *Journal of the Audio Engineering Society*, 42(1/2), 1994.
- Macpherson, E.A. A Computer Model of Binaural Localization for Stereo Imaging Measurement. Presented at the 87th Audio Engineering Society Convention, New York, USA, 1989.

- Madisetti, V.K. & Williams, D. *The Digital Signal Processing Handbook* 1st ed., CRC, 1997.
- Makita, Y. On the Directional Localisation of Sound in the Stereophonic Sound Field. *E.B.U. Review*, 73(Part A - Technical), pp.1536-1539, 1962
- Makous, J.C. & Middlebrooks, J.C. Two-dimensional Sound Localization by Human Listeners. *The Journal of the Acoustical Society of America*, 87(5), pp.2188-2200, 1989.
- Malham, D.G. Computer Control of Ambisonic Soundfields. Presented at the 82nd Audio Engineering Society Convention, London, UK, 1987.
- Malham, D.G. Experience with large area 3-D Ambisonic Sound Systems. Proceedings of the Institute of Acoustics, 14(5), pp. 209-216, 1992.
- Malham, D.G. Higher Order Ambisonic Systems for the Spatialisation of Sound. Proceedings of the International Computer Music Conference, Beijing, China, pp. 484-487, 1999.
- Mapp, P. Psychoacoustics in Sound Reinforcement and PA System Design. Presented at the 22nd UK Audio Engineering Society Conference, *Illusions in Sound: the application of psychoacoustics to audio*. Cambridge, UK, 2007.
- Marler, R.T. *A Study of Multi-objective Optimization Methods for Engineering Applications*. PhD. The University of Iowa, 2005.
- Marler, R.T. & Arora, J.S. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6), pp.369-395, 2004.
- Martin, G., Woszczyk, W., Corey, J. & Quesnel, R., Sound Source Localization in a Five-Channel Surround Sound Reproduction System. Presented at the 107th Audio Engineering Society Convention New York, USA, 1999.
- Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1998.
- Mills, A.W. Auditory Localization. In J. V. Tobias, ed. *Foundations of Modern Auditory Theory*. New York: Academic Press, p. 651. 1972. Available at: <http://adsabs.harvard.edu/abs/1972JSV....25..651T> [Accessed January 12, 2009].

- Mills, A.W. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, 30, pp.237-246, 1958.
- Misevicius, A. A tabu search algorithm for the quadratic assignment problem. *Comput. Optim. Appl.*, 30(1), pp.95-111, 2005.
- Moller, H., Sorensen, M., Jensen, C. B. & Hammershoi, D. Binaural Technique: Do we need Individual Recordings? *Journal of the Audio Engineering Society*, 44(6), pp.451-469, 1996.
- Moore, B.C., Oldfield, S.R. & Dooley, G.J. Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. *The Journal of the Acoustical Society of America*, 85(2), 820-36, 1989.
- Moore, B. *An Introduction to the Psychology of Hearing*, Academic Press, 2003.
- Moore, B., Oldfield, S.R. & Dooley, G.J. Detection and Discrimination of Spectral Peaks and Notches at 1 kHz and 8 kHz. *The Journal of the Acoustical Society of America*, 85(2), 820-836, 1989.
- Moore, D. & Wakefield, J.P. Exploiting Human Spatial Resolution in Surround Sound Decoder Design. Presented at the 125th Audio Engineering Society Convention, San Francisco, California, USA, 2008.
- Moore, D. & Wakefield, J.P. The Design and Detailed Analysis of First Order Ambisonic Decoders for the ITU layout. Presented at the 122th Audio Engineering Society Convention, Vienna, Austria, 2007.
- Moore, D. & Wakefield, J.P. The Design of Ambisonic Decoders for the ITU 5.1 Layout with Even Performance Characteristics. Presented at the 124th Audio Engineering Society Convention, Amsterdam, The Netherlands, 2008.
- Moore, D. & Wakefield, J.P. The Design of Improved First Order Ambisonic Decoders by the Application of Range-Removal and Importance in a Heuristic Search Algorithm. Presented at the 31st International Audio Engineering Society Conference, London, UK, 2007.
- Moore, D. & Wakefield, J.P. The Potential of High Performance Computing in Audio Engineering. Presented at the 126th Audio Engineering Society Convention, Munich, Germany, 2009.
- Moreau, S., Daniel, J. & Bertet, S. 3D Sound Field Recording with Higher Order

- Ambisonics - Objective Measurements and Validation of Spherical Microphone. Presented at the 120th Audio Engineering Society Convention, Paris, France, 2006.
- Muraoka, T. & Nakazato, T. Examination of Multichannel Sound-Field Recomposition Utilizing Frequency-Dependent Interaural Cross Correlation (FIACC). *Journal of the Audio Engineering Society*, 55(4), pp.236-256, 2007.
- Nam, J., Abel, J.S. & Smith, J.O. A Method for Estimating Interaural Time Difference for Binaural Synthesis. Presented at the 125th Audio Engineering Society Convention San Francisco, USA, 2008.
- Neukom, M. Decoding Second Order Ambisonics to 5.1 Surround Systems. Presented at the 121st Audio Engineering Society Convention, San Francisco, USA, 2006.
- Noble, W. & Gates, A. Accuracy, Latency, and Listener-Search Behaviour in Localization in the Horizontal and Vertical Planes. *The Journal of the Acoustical Society of America*, 78(6), pp.2005-2012, 1985.
- Nowicki, E. & Smutnicki, C. A fast taboo search algorithm for the job shop problem. *Manage. Sci.*, 42(6), pp.797-813, 1996.
- Oldfield, S.R. & Parker, S.P.A. Acuity of Sound Localisation: a Topography of Auditory Space. 1. Normal Hearing Conditions. *Perception*, 13(5), pp.581-600, 1984.
- Ozcan, K., Busbridge, S.C. & Fryer, P.A. Determination of the Relative Hierarchy of Audible Cues in Conflict. Presented at the 112th Audio Engineering Society Convention, Munich, Germany, 2002.
- Ozcan, K., Busbridge, S. C., Fryer, P., Geaves, G. P. & Moore, J. P. The Significance of Phase as an Auditory Cue. Presented at the 114th Audio Engineering Society Convention, Amsterdam, The Netherlands, 2003.
- Painter, T. & Spanias, A. Perceptual coding of digital audio. *Proceedings of the IEEE*, 88(4), pp.451-515, 2000.
- Patterson, R.D., Allerhand, M.H. & Giguere, C. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, 98(4), pp.1890-1894, 1995.
- Pernaux, J., Boussard, P. & Jot, J. Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis. In the Proceedings of the

- Workshop on Digital Audio Effects, Barcelona, Spain, 1998.
- Pocock, M. A Computer Model of Binaural Localization. Presented at the 72th Audio Engineering Society Convention, California, USA, 1982.
- Poletti, M.A. Robust Two-Dimensional Surround Sound Reproduction for Nonuniform Loudspeaker Layouts. *Journal of the Audio Engineering Society*, 55(7/8), pp.598-610, 2007.
- Poletti, M.A. Three-Dimensional Surround Sound Systems Based on Spherical Harmonics. *Journal of the Audio Engineering Society*, 53(11), pp.1004-1025, 2005.
- Pulkki, V. & Hirvonen, T. Localization of virtual sources in multichannel audio reproduction. *Speech and Audio Processing, IEEE Transactions on*, 13(1), pp.105-119, 2005.
- Pulkki, V. Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning. *Journal of the Audio Engineering Society*, 49(9), pp.753-767, 2001.
- Pulkki, V. Spatial Sound Reproduction with Directional Audio Coding. *Journal of the Audio Engineering Society*, 55(6), pp.503-516, 2007.
- Pulkki, V. & Karjalainen, M. Localisation of Amplitude Panned Virtual Sources 1: Stereophonic Panning. *Journal of the Audio Engineering Society*, 49(9), pp.739-752, 2001.
- Pulkki, V., Karjalainen, M. & Huopaniemi, J. Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model. *Journal of the Audio Engineering Society*, 47(4), 203-217, 1999.
- Rayleigh, L. On Our Perception of Sound Direction. *Philosophical Magazine*, 13, pp.214-232, 1907.
- Rossi-doria, O., Sampels, M., Birattari, M., Chiarandini, M., Dorigo, M., Gambardella, L., Knowles, J., Manfrin, M., Mastrolilli M., Paechter, B., Paquete, L. & Stutzle, T. A comparison of the performance of different metaheuristics on the timetabling problem. *In the Proceedings of the 4th International Conference on Practice and Theory of Automated Timetabling (PATAT 2002)*, 2740, pp.329-351, 2002.
- Rumsey, F. *Spatial Audio*, Focal Press, 2001.

- Saberi, K., Dostel, L., Sadralodabai, T. & Perrott D. R. Minimum Audible Angles for Horizontal, Vertical and Oblique Orientations: Lateral and Dorsal Planes. *Acustica*, 75, 1991.
- Sæbø, A. Implementation of Transaural Systems in Software on a PC. Presented at the 105th Audio Engineering Society Convention, San Francisco, USA, 1998.
- Saleh, H.A. & Dare, P. Effective Heuristics for the GPS Survey Network of Malta: Simulated Annealing and Tabu Search Techniques. *Journal of Heuristics*, 7(6), pp.533-549, 2001.
- Segal, B., Robertson, L., Gagliardi, F. & Carminati, F. Grid computing: the European Data Grid Project. In *Nuclear Science Symposium Conference Record, 2000 IEEE*. p. 2/1 vol.1, 2000.
- Slaney, M., Auditory Toolbox. Available at:
<http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/> [Accessed July 8, 2009].
- Smith, S. W. The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publications, 1998.
- Sontacchi, A., Noisternig, M., Majdak, P. & Holdrich R. An Objective Model of Localisation in Binaural Sound Reproduction Systems. Presented at the 21st International Audio Engineering Society Conference, St Petersburg, Russia, 2002.
- Stevens, S.S. & Newman, E.B. The Localization of Actual Sources of Sound. *The American Journal of Psychology*, 48(2), pp.297-306, 1936.
- Strybel, T.Z. & Fujimoto, K. Minimum Audible Angles in the Horizontal and Vertical Planes: Effects of Stimulus Onset Asynchrony and Burst Duration. *The Journal of the Acoustical Society of America*, 108(6), pp.3092-3095, 2000.
- Ternstrom, B. Radio in 5.1 - The True Experience. Presented at the 24th International Audio Engineering Society Conference, Banff, Canada, 2003.
- Theile, G. & Plenge, G. Localization of Lateral Phantom Sources. *Journal of the Audio Engineering Society*, 25(4), pp.196-200, 1977.
- Theile, G. HDTV Sound Systems: How many Channels? Presented at the 10th International Audio Engineering Society Conference, London, UK, 1991.

- Theile, G. The New Sound Format "3/2-Stereo". Presented at the 10th International Audio Engineering Society Conference, Berlin, Germany, 1993.
- Theiß, B. & Hawksford, M. Binaural Model-Based Measurements of Phantom Images. Presented at the 105th Audio Engineering Society Convention, San Francisco, USA, 1998.
- Thurlow, W.R. & Runge, P.S. Effect of Induced Head Movements on Localization of Direction of Sounds. *The Journal of the Acoustical Society of America*, 42(2), pp.480-488, 1967.
- Vanderkooy, J. & Lipshitz, S.P. Anomalies of Wavefront Reconstruction in Stereo and Surround Sound Reproduction. Presented at the 83rd Audio Engineering Society Convention, New York, USA, 1987.
- Wallach, H., Newman, E.B. & Rosenzweig, M.R. The Precedence Effect in Sound Localization. *The Journal of the Acoustical Society of America*, 21(4), pp.468, 1949.
- Ward, D. & Abhayapala, T. Reproduction of a plane-wave sound field using an array of loudspeakers. *Speech and Audio Processing, IEEE Transactions on*, 9(6), pp.697-707, 2001.
- Webster, R. Useful AI tools-a review of heuristic search methods. *Potentials, IEEE*, 10(3), pp.51-54, 1991.
- Wehner, M., Olikar, L. & Shalf, J. Towards Ultra-High Resolution Models of Climate and Weather. *The International Journal of High Performance Computing Applications*, 22(2), pp.149-165, 2008.
- Wenzel, E.M., Arruda, M., Kistler, D. J. & Wightman, F. L., 1993. Localization using Non-Individualized Head-Related Transfer Functions. *The Journal of the Acoustical Society of America*, 94(1), pp.111-123.
- Westhead, D.R., Clark, D.E. & Murray, C.W. A comparison of heuristic search algorithms for molecular docking. *Journal of Computer-Aided Molecular Design*, 11(3), pp.209-228, 1997.
- Wiggins, B. *An Investigation into the real-time Manipulation and Control of Three-Dimensional Sound Fields*. PhD. University of Derby, 2004.

- Wiggins, B. Has Ambisonics Finally Come of Age? In the Proceedings of the Institute of Acoustics, 30(6), 2008.
- Wiggins, B. The Generation of Panning Laws for Irregular Speaker Arrays using Heuristic Methods. Presented at the 31st International Audio Engineering Society Conference, London, UK, 2007.
- Wiggins, B., Paterson-Stephens, I., Lowndes, V. & Berry, S. The Design and Optimisation of Surround Sound Decoders Using Heuristic Methods. Proceedings of UKSim 2003, Conference of the UK Simulation Society p.106-114, 2003.
- Wiggins, B., Paterson-Stephens, I. & Schillebeeckx, P. The analysis of multi-channel sound reproduction algorithms using HRTF data. Presented at the 19th International Audio Engineering Society Conference, Germany, 2001.
- Wightman, F.L. & Kistler, D.J. The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization. *The Journal of the Acoustical Society of America*, 91(3), pp.1648-1661, 1992.
- Wightman, F.L. & Kistler, D.J. Resolution of front--back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5), pp.2841-2853, 1999.
- Wright, D., Hebrank, J.H. & Wilson, B. Pinna Reflections as Cues for Localization. *The Journal of the Acoustical Society of America*, 56(3), pp.957-962, 1974.
- Zahorik, P. Perceptual Recalibration in Human Sound Localization: Learning to Remediate Front-Back Reversals. *The Journal of the Acoustical Society of America*, 120(1), pp.343-359, 2006.
- Zotkin, D., Hwang, J., Duraiswaini, R. & Davis, L. S. HRTF personalization using anthropometric measurements. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. pp. 157-160, 2003.